

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 31, 2002, 14:34:58 ; Search time 14.65 Seconds

(without alignments)
1188,768 Million cell updates/sec

Title: US-09-905-088a-245

Perfect score: 3732

Sequence: 1 MRLVAPLLAMVAGATATV.....RKLPSSSECTLLPPLSONS 713

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgn2_6/prodata/2/1aa/5A.COMB.pep:*

2: /cgn2_6/prodata/2/1aa/5B.COMB.pep:*

3: /cgn2_6/prodata/2/1aa/6A.COMB.pep:*

4: /cgn2_6/prodata/2/1aa/6B.COMB.pep:*

5: /cgn2_6/prodata/2/1aa/6C.COMB.pep:*

6: /cgn2_6/prodata/2/1aa/6D.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Query Match	length	DB ID	Description
1	1656.5	44.4	708	4	US-09-131-648-2
2	442	11.8	1091	3	US-08-986-485-5
3	408	10.9	1101	3	US-08-986-485-2
4	407.5	10.9	605	1	US-08-190-802A-49
5	407.5	10.9	605	4	US-08-477-346-49
6	407.5	10.9	605	4	US-08-473-089-49
7	402	10.8	673	4	US-09-063-950-2
8	397.5	10.7	605	4	US-09-063-950-5
9	355.5	9.5	1523	4	US-09-182-024A-2
10	344	9.2	603	1	US-08-190-802A-50
11	344	9.2	603	4	US-08-477-346-50
12	344	9.2	603	4	US-08-473-089-50
13	331	8.9	1525	3	US-09-191-647-2
14	331	8.9	1525	4	US-09-191-647-2
15	331	8.9	1525	4	US-09-191-647-2
16	324	8.7	1480	3	US-09-540-153-2
17	324	8.7	1480	4	US-09-540-153-2
18	324	8.7	1480	4	US-09-540-153-7
19	324	8.7	1480	4	US-09-182-024A-5
20	324	8.7	1480	4	PCT-US91-09055-2
21	321	8.6	560	3	US-08-592-500-2
22	321	8.6	560	3	US-08-195-006-2
23	321	8.6	560	4	US-09-063-950-4
24	321	8.6	560	4	PCT-US94-07644A-2
25	316.5	8.5	353	3	US-08-986-485-6
26	296.5	7.9	649	4	US-09-188-930-305
27	274	7.3	1112	4	US-09-353-585-2

28	273	7.3	1112	4	US-09-353-585-3	Sequence 3, Appl1
29	269	7.2	353	4	US-09-180-439-8	Patent No. 5340934
30	263	7.0	1016	4	US-09-442-063A-45	Sequence 48, Appl1
31	261	7.0	282	1	US-08-442-063A-48	Sequence 4, Appl1
32	261	7.0	337	1	US-08-442-063A-27	Sequence 27, Appl1
33	261	7.0	337	1	US-08-442-063A-27	Sequence 2, Appl1
34	261	7.0	342	1	US-08-612-916-2	Sequence 2, Appl1
35	261	7.0	342	5	US-08-612-916-2	Sequence 4, Appl1
36	261	7.0	342	5	US-08-612-916-2	Sequence 4, Appl1
37	261	7.0	353	1	US-08-458-834-4	Sequence 4, Appl1
38	260	7.0	353	1	US-08-458-834-4	Sequence 4, Appl1
39	260	7.0	353	1	US-08-458-834-4	Sequence 4, Appl1
40	260	7.0	353	1	US-08-458-834-4	Sequence 4, Appl1
41	250	6.7	799	4	US-09-180-439-6	Sequence 6, Appl1
42	245.5	6.6	368	1	US-08-303-238-3	Sequence 3, Appl1
43	245.5	6.6	368	1	US-08-458-834-3	Sequence 3, Appl1
44	237.5	6.4	368	6	US-09-353-585-2	Patent No. 5340934
45	237.5	6.4	746	5	PCT-US95-10509-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-09-131-648-2
Sequence 2, Application US/09131648
Patent No. 6168920
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Yue, Henry
APPLICANT: Corley, Neil C.
APPLICANT: Guejter, Karl J.
TITLE OF INVENTION: EXTRACELLULAR ADHESIVE PROTEINS
FILE REFERENCE: PE-0576 US
CURRENT APPLICATION NUMBER: US/09/131,648
CURRENT FILING DATE: 1998-08-10
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PERL Program
SEQ ID NO 2
LENGTH: 708
TYPE: PRT
ORGANISM: Homo sapiens
FEATURES:
OTHER INFORMATION: 2687731
US-09-131-648-2

Query Match 44.4%; Score 1656.5; DB 4; Length 708;
Best Local Similarity 47.2%; Pred. No. 2.6e-144;
Matches 333; Conservative 109; Mismatches 217; Indels 47; Gaps 5;

QY	17	TATVPVPMHVCPPQACQIRPWYTPRSSYREATTVCNDIFLAVPALPAGTQILL	76
DB	17	TTTVAQVDRKVKPCPLCTCEIRPWFPTPSIYMASVTDCNDIGLTPPALPANTQILL	76
QY	77	QNSIYRVDSLEGLANTLETDLSSQNSFPDARDCEFHLPQLLSLHLENDITLEDHS	136
DB	77	QTNIAKIEYS-TFPVNLTLGLDLSQNMSSVINWKKMPQLSYLEENKLTLPKPC	135
QY	137	FAGIASLOELYNHOLYRIAPRAESGSLRLHLNSMLRAIDSWFEMLPNTEILLI	196
DB	136	LSEISNLOELYNHMLSTISPGAFGLHNLRLHLNSRLQMSINSMFALPNEILLI	195
QY	197	GCNNVAIILNDNFRPLNLSLVLGMMIAREISVALLEGISLESSEFYNDOLANVRA	256
DB	196	GENIIRIKDMFKPLINLSVLINGIMLTIPNALVLENTLSISFYNNRLITVPIVA	255
QY	257	LEQVPGKFLDLKNKPLORVGPDFANMLHKLGLGNNNEELVSDKFAVLNPELTLD	316
DB	256	LQKVVNLFKFLDNKMPINIRIRGDSNNMLHKLGLGNNNEELVSDKFAVLNPELT	315
QY	317	ITNPRLSFIHRAHVFPMETLMLNNLSALHOOVYESI.PMLQVJGAGNPTRCDV	376

Dh	316	ATNNRLSYIHMANAFPELRKLSLMSLMSALMYIGTIESLPMLEKSHISHPRLRCOV	375
Oy	377	IMANANSTGRVRFTEPOSTLCAEPDLDLORLPVEEVPFRRMDDCLPLISRFSPSLQVA	436
Dh	376	IRMANNNNTNTRFEPDPSLCVDFPEEGGVNBOVHFRRMBEDCLPLIAESPSPLNVE	435
Oy	437	SGEHWYLHCRAALAEPEBELIYWTVPAGLRLPALPAGRGYRVYPEGCTELRRVYAEAGLYT	496
Dh	436	AGSYVSEFCRATAPDPELYTWTPTEGCKLLPNTLTDRKYVHSEDTLINVYPRKGGIYT	495
Oy	497	CYANOLVADPTKTVSVYVVGRRALLDGRDEGQGLTKVDETFRHYLLDSWTPPEPNTVSTL	556
Dh	496	CIATLRLVADLKSVMKLVGSGPPO---DNNGSLNIRKIRPDQANSVLVSMWASSKILTKSSV	552
Oy	557	TWSSASSLRGGCATALARLPGRTHSYNTRLLDQATEWACIQVAFADATOLACAWATK	616
Dh	553	KTATATVKTENSHAQSAKRLPSVRYKNTLTPLNSTEYKCIDPTIYOKNRKRCVNTYK	612
Oy	617	-----EATSCRRALCDRGRGLATLALAVLLLAAGLAHLTGQPRKVGGR-	663
Dh	613	GLHPDQREYKNNNTTLTMACLGGLGIGIYICISLT-----SPENMDGGHS	660
Oy	664	-----PLPAPAAEWGMSAPSVRYVAPLTPMN	691
Dh	661	TYRNTLQKPTPLGELYPLPLINLWDACKERKSTLAKKATVYIGLPTN	706

```

US-08-966-485-5          ? US-08-966-485-5
                            ? RESULT 2
? Sequence 5, Application US/08986485      ?
? Patent No. 6046030                      ?
? GENERAL INFORMATION:                     ?
? APPLICANT: WU, SHUIJIAN                 ?
? APPLICANT: SWEET, RAYMOND              ?
? APPLICANT: TUNGH, ALEMSPEED            ?
? TITLE OF INVENTION: A HUMAN LIG-I HOMOLOG (HLIS-1)?
? NUMBER OF SEQUENCES: 8                  ?
? CORRESPONDENCE ADDRESS:                ?
? ADDRESSEE: RATNER & PRESTIA           ?
? STREET: P.O. BOX 980                   ?
? CITY: VALLEY FORGE                    ?
? STATE: PA                              ?
? COUNTRY: USA                           ?
? ZIP: 19482                             ?
? COMPUTER READABLE FORM:               ?
? MEDIUM TYPE: diskette                 ?
? COMPUTER: IBM Compatible               ?
? OPERATING SYSTEM: DOS                  ?
? SOFTWARE: FASTSPD for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/966,485
FILING DATE: 08-DEC-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/059,448
FILING DATE: 22-SEP-1997
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GH-70264
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEX: 610-407-0701
INFORMATION FOR SEQ ID NO.: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1091 amino acids
TYPE: amino acid
STRANDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
```

Query Match 11.8%; Score 442; DB 3; length 1091;
 Best Local Similarity 23.2%; Pred. No. 1.1e-31;
 Matches 169; Conservative 106; Mismatches 255; Indels 198; Gaps 22.

```

01 37
02 5 VILLAW- - - - - AGATATVPVPHVYCPPOCAC- - - - -
03 14 IABRLIMLILLILLIONPESAGAOAB- - - - - RACCAACTCGAGNSLDCSGNGLATYPRDL
04 38 RPYWYTRS- - - SYEATVY- - - - - CNDLEIVVPALPA- - - GTOTILLDS 78
05 69 PSM- - - TRSNLSTYRRSEIDSAFEDLTINOEVYLSNNEITAPISLGAASGVVSLDLOH 120
06 79 NSTVRDSELTJANTITELDSONSFPDADCCFALPOLLSTLEBNOITLYDEHSPA 130
07 127 NKLLVDSGOLKSYLSLEVLDLSSNNITEISSCFNGRIEMLWLSNRSTIEGAD 180
08 139 GLA- - - STOEIYLNHQLRIAPRAFSGLSNLRJLHNSNLRAIDSWEFEMJPLEITMIG 194
09 187 GJRSRLTLTRSKRRTIQLEKAFK- - - LPRLOQDLKNNRIRILBGLTPOGDSLEVLRLQ
10 198 GKWDALIDNFRPLANLSVLAQMLREISDYLEGSLQSLSEISFYDQOLARVR- - - 254
11 246 RNNISLTDGAFWGLSNHVLHLEYSLVEVNSGSLYGLTALHOLHLSNNSIRIOGOM 305
12 255 - - - - - RALEOVPGLKEITDLKNRPLQVNGRDPANMLIKELGLN 295
13 306 SFQOKIHELITSFNNLTRIDEBLSLSSILRLSHNLSIHABGAFKCKSLRVLDD 365
14 294 NMEELVSD- - - KFLVNLPELTKEDITNNPLSFIHPRAFHMLPOMETLMLNNALSALH 355
15 366 HNEISOTLEDTSGLFTGLDNLSKTLFEGN- - - KISVAKRAFSGLESDEHMLGNATRSQ 422
16 352 QCVTESLNOEVLGHNPLRDCVYTR- - - - - ANAT- - - - - GTRVRFI 390
17 425 FQAFKAKNKEVLISESPFLDCOLKPLRMYLGMKLOAFYATCAHPSLKSGSITSV 480
18 391 EPOSLCAEP- - - - - DL 408
19 485 LPDSVFCDDPKPKQIITQFETTMAYVAGDIRTCSAASSSPMTFAMKNDVYLANDM 544
20 404 QR- - - - - LPRVRF- - - - - REMDHCPLIS 422
21 545 ENRAVRAOGEVWEYTTILHNRHTFEHNRGOCITYNHGSYTSKANTLYNVLP- - - S 602
22 426 PRSPFSLVAGSGSWVLCGRALAREPELYWTAGTLEPRVAGRGVVR- - - ESTLE 484
23 603 FRTHPDIAIRGTYARLECATGHPNQIAMOKDGTDF- - - PAAREBRRHVMYDDOVFI 664
24 485 RYVTEEGELTYCAONLVC- - - ADTKVYVAVVRALALOGRE- - - GQLELPAOENR 538
25 662 JYVKRIDDKGVYSCTAONSAGSVSANATITVLETPSLAVPLEDRVYVGTETVAPOCKATGS 721
26 539 YHILSNW 546
27 722 PPRITWL 729
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
81
```

Sat. Aug 31 14:55:21 2002

us-09-905-088a-245.rai

Page 3

```

1 STATE: PA
2 COUNTRY: USA
3 ZIP: 19482
4 COMPUTER READABLE FORM:
5 MEDIUM TYPE: Diskette
6 COMPUTER: IBM Compatible
7 OPERATING SYSTEM: DOS
8 SOFTWARE: FastSeq for Windows Version 2.0
9 CURRENT APPLICATION DATA:
10 APPLICATION NUMBER: US/08/986,485
11 FILING DATE: 08-DEC-1997
12 CLASSIFICATION:
13 PRIOR APPLICATION DATA:
14 APPLICATION NUMBER: 60/059,448
15 FILING DATE: 22-SEP-1997
16 ATTORNEY/AGENT INFORMATION:
17 NAME: PRESTIA, PAUL F
18 REGISTRATION NUMBER: 23,031
19 REFERENCE/DOCKET NUMBER: CH-70264
20 TELECOMMUNICATION INFORMATION:
21 TELEPHONE: 610-407-0700
22 TELEFAX: 610-407-0701
23 TELEX: 846189
24 INFORMATION FOR SEQ ID NO: 2:
25 SEQUENCE CHARACTERISTICS:
26 LENGTH: 1101 amino acids
27 TYPE: amino acid
28 STRANDEDNESS: single
29 TOPOLOGY: linear
30 MOLECULE TYPE: protein
31 OS-08-986-485-2

```

Query Match	10.9%	Score 408	DB 3	Length 1101
Best Local Similarity	21.8%	Poed 1.5e-28		
Matches 185	Conservative 124	Mismatches 277	Indels 264	Gaps 30
QY	8	LLIMV-----	RGATATVPPVPMHWCPQCAOIRPWTPRASSREATVVCN	56
Db	18	LLILLMLLELRLEPYTAAGPRA-----	PCAAATOCAGPCTCAGDS-----	100G 62
QY	57	DLPLTAVPPLPAGTQTLLOLSNVIYVDSGLATNLT-----		97
Db	63	GRGLALPGDLPWTRSLNTSYNKKLAETDPAGFPDLPNIQVYLNHETAVASLGASS		122
QY	98	-----	LDLSQSPDARDDDPALPOLLSLHLEENOL	129
Db	123	QYVALTFQOQONRSIDGSOQKAYLSLEVLDLNTLNITTEVRNNYTFPHGPRIKELNTLAGRI		182
QY	130	PRLEDSHFAGLA-STQELLYNHOLRYTAPRAFSGLSNLRILHNSNLTAIDRSMFEL		188
Db	183	GTLELGAFPDLSLSILLTRLSKRRTIQLPVPARF-LPRLLQDLDNRNRINRLIEGLTQGL		241
QY	189	PULPLMIGGNVYDALIDMNPRLANRSTVLAQMRPEISYALAEGLSLEISSTYQNO		248
Db	242	NSLEVLKLONNISKLTDGAFWGLSGMHWLHLEDSIVAYNGSLYGLTALDHLSNNS		301
QY	249	LAAPRR-----	ALQVYGLKFLDILKNNLDRVGGPFAAM	284
Db	302	IAIRHRGWFCQKALHLYLVSFNMLRLRDESLAEISLSYLRHSNISHLNKGAFKL		361
QY	285	LHKLELDLNNBELYSID----KFALYNLPELTALDTINNPALSTIRPAFHHLPOMETL		340
Db	362	RSRLKLDLDNNELSGTIDTSGAFSGLEGR-SKLTLPFG-KIKSVAKRAFSGLSGLHLL		419
QY	341	MLNNALSLAHQQTVESLPNLQEVYLGHNPRICDCVIRW-----ANAT----		383
Db	420	NLGGALATRSVQDPAFVKKRNKLKELHLSSTPFLCDQQLKMLPWLIGMLQAFYATACAP		479
QY	384	-----GTRAFIEPOSTLC-----AEP-----		400
Db	480	ESLAGQSIYFSPESVCDQFLPQITPQPTTAAVYKDIRFTCSAASSSSPMFAWK		539

```

QY      401 -----POLCR-----LPVREVPF-----R 413
Db      540 KONEVLTAADKEKFNVAHAODEVEWETTLLHNRQTATGEGHCPRYOCTNNHGSTYSKA 599
QY      415 EMTDIDCLTPISPSPPSPSLQVAGSGSVLHCRALEAPPEETIYWTPAGRLPAHAGRRY 474
Db      600 RLTVAVLP--SEKPRTHTITRIPTVTTLVARLECAATGTGNPRLAMOKOGGRDF-PAARERM 656
QY      475 RYVP-EGLTELRATYAEEAGLTGVACNTVG--ADTKTIVSYVGRALDPGRDE----GO 527
Db      657 HMPEDDDVFPIVDVKLDQAGYSCIAQNSAGSLSANMLTYLETPLPVPLDRVSWSGE 716
QY      528 GELINVOCHTHYHLHSWTNPRTNVSTNMLSASLSLGOCATLARLRGTHSNYT-- 585
Db      717 TVALCKRKTG-----NPF-----PRIW-----FKGDHPISL-----TERNHLPD 752
QY      586 -RLU-----CATPYMNCLO-VAAADAHPOLAQMANKKETTSCHRALDRRGILAI 634
Db      753 NQLLVQVVNAVEDNGKRYTCGMENSLGTERAHSLSYL-----PALGC-KKDSTYGIFTI 806
QY      635 LALAFLILAA 644
Db      807 AVASSIVLTS 816

RESULT          4
US-08-190-802A-49
Sequence 49, Application US/08190802A
Patent No. 5519003
GENERAL INFORMATION:
APPLICANT: Mochly-Rosen, Daria
APPLICANT: Ron, Dorit
TITLE OF INVENTION: WD-40 - Derived peptides and Uses
TITLE OR INVENTION: Theroef
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: P.O. Box 60850
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-0850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/190,802A
FILING DATE: 01-FEB-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 8600-0139
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO.: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 605 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: Insulin-like growth factor binding
INDIVIDUAL ISOLATE: protein complex, Fig. 32
US-08-190-802A-49
```

Query Match 10.98; Score 407.5; DB 1; Length 605;

Best Local Similarity 25.2%; Pred. No. 6.1e-29;
Matches 152; Conservative 72; Mismatches 173;
Indels 205; Gaps 16

OY	3	LIVAEPLILAWVA	-----	GATATVPVVMHNVPOCAQIRWYBRSSY	-----	REATT	52
Db	8	LALALLLSVVALGPRSLTEGADPCPCBAGCAPACACVC	-----	SYDDADELS	57		
OY	53	VQONDLFTVAPPALPACGQTLLLOSNSTYRV	-----	DOSLIG	91		
Db	58	VFCSSRNLTFLPQVPGVGTQALMDGNLSSVPAPAFQWLSIGFLMQGGQGLSUEPOA	117	117			
OY	92	-----LANLEEDLOSNSFSQADRCDFHALPOLSLHREMOULRE	-----	133			
Db	118	LGLLENCHLERNOLNRSLATGFAHTPALASIGSNRSLREDDUFLRGLOSLMDNL	177	177			
OY	134	-----DHSPFGLASIOEYLXNHOLXIRAPAFGTSNULRHLNLLRLAIDSRW	184	184			
Db	178	GMSLSIAPRPAARGLGSLAREYLAGNRLAYDLPALFSGLEHRELOLSNNALRIKANY	237	237			
OY	185	FEMLEPNFILMIGGNVYDI	-----	LDN	208		
Db	238	FQJDRLOKLYIDBRNLIAAVAGAFGLIKALMHDLSHNRKACGLETFPGLIGLRURL	297	297			
OY	209	-----PRPLANRSYLACMNLREISDYALBGLOSLESFSFYDOLARY	252	252			
Db	298	SBNALIASLRPRTDRHLEELQGLQHNRIQLAESNFSFGQLEVLTLIDHNOLEKAGA	357	357			
OY	253	-----PRRLAEVPOGLATLDLNKRLQVRYGQDFANMLHLKELG	292	292			
Db	358	FLGLTNVAVMNLGNCRLRLPQVYPRGKGLSHLHESGCAGRIPTPTFGSLGRRLFL	417	417			
OY	293	NNKNEELVIDKFAVLNLPETLRKLDITNNPLUSTIFPRAHNLPOMETLMLNNALSAL	350	350			
Db	418	KD-NGLVETIEOSKMGJALTELDLJTSN-QUTHLPRILPQGLKLEYLLLSRRRLAEIPA	479	479			
OY	351	-----HOQVIESLPN	-----	LQEVGLHG	368		
Db	476	DALGELQAFAPLQVSHNRLEALPNSLAPLARGRLYLSLRNNSLRTPTPQPGGLERMLLEG	535	535			
OY	369	NPIRCDCVIR-----WANATGTVR-RTLEP	-----	OSTICAEPPDLORLPVR	409		
Db	536	NPWDCGCPFLKALRDFALONPSAVPREVQALCEBGDDCQPRAYVYNNITCASPREVYGLDLR	595	595			
OY	410	EV	411				
Db	596	DL	597				

RESULT 5
US-08-477-346-49
; Sequence 49, Application US/08477346

GENERAL INFORMATION:
APPLICANT: Mochly-Rosen, Daria
APPLICANT: Ron, Dorit
TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
TITLE OF INVENTION: thereof
NUMBER OF INVENTIONS: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 2000 Pennsylvania Avenue, NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,346
FILING DATE: 07-JUN-1995

```

1 CLASSIFICATION: 514
2 PRIOR APPLICATION DATA:
3 APPLICATION NUMBER: 08/487,072
4 FILING DATE: 07-JUN-1995
5 ATTORNEY/AGENT INFORMATION:
6 NAME: MURASHIGE, KANE H.
7 REGISTRATION NUMBER: 29,959
8 REFERENCE/DOCKET NUMBER: 2550-0025.20
9 TELECOMMUNICATION INFORMATION:
10 TELEPHONE: (202) 887-1500
11 TELEFAX: (202) 887-0763
12 INFORMATION FOR SEQ. ID. NO.: 49:
13 SEQUENCE CHARACTERISTICS:
14 LENGTH: 605 amino acids
15 TYPE: amino acid
16 TOPOLOGY: unknown
17 MOLECULE TYPE: Protein
18 HYPOTHEICAL: NO
19 ANTI-SENSE: NO
20 ORIGINAL SOURCE:
21 INDIVIDUAL ISOLATE: Insulin-like growth factor binding
22 INDIVIDUAL ISOLATE: protein complex, Fig. 32
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997

```

Query Match	10.98;	Score 407.5;	DB 4;	Length 605;
Best Local Similarity	25.28;	Pred. No. 6.1e-29;		
Matches 152;	Conservative	72;	Mismatches 173;	Indels 205;
				Gaps 16;

QY	3	LIVALLALLAWA	-----	GATATVPVVMHNYCPPOCAQIIRWYPRSSY	-----	REATT	52
		1	1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1		
Dd	8	LALALLLSWALGCPRESLEGADPCTGEBAGPACPACVC	-----	SYDDADELS	-----		57
QY	53	VOCNDLFTTAVPAPLPAQTOTLLLOGSIYRV	-----	DOSELGY	-----		91
		1	1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1		
Dd	58	VECSRNITRTPGVGGTQALMTLGDNNLTSSVPAFQNTSLGFLMLGGGSLBPQA	1177				
QY	92	---LANTJELDSONSFSPARCDENALPOLSLHNEEQUITRE	-----				133
		1	1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1		
Dd	118	LIGTENLOCHTERNOJRLSLATGCTPAHPALASIGSNRSLREDELFGGLSLMDNL	1777				
QY	134	-----DHSEFGLASIOELYLNNHOLRIAPRFGSLMLRLNLSNLAIDSRW	184				
		1	1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1		
Dd	178	GMNSLAVLPDAERGLGSLRELYLGNRLVLAQPAFSGSLAETRELDLSRNALRAIKANV	233				
QY	185	FEMLNEIEMIGSNVDA	-----	LDKN	-----		208
		1	1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1		
Dd	238	FQJPERQKTYIDRNLIAVAPGAFGLKALRMJLDSHNVAGLLEDTPTGGLGLRVRL	297				
QY	209	-----PRPLANRSLVLAGNNEISDVLEISLSEFYONOLAV	252				
		1	1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1		
Dd	298	SHNALASLRPTKRDHLFELELOLGNHRITQALNERFEBLEQLEVLILDHNOJOEYKXGA	357				
QY	253	-----PRALDEPOLKTYDANKPRLOVQPDYNNHILKEGL	292				
		1	1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1		
Dd	358	FLGLTNVANNLSGNOCLRPLQVYFRGIGLHSLHLESCGLRTRPTTIGLSGLRFL	411				
QY	293	NNMEIYSIDKRALVNLPELTCLDITNNRPLSTIYHRAFNHLPOEMETIMNNALSL	350				
		1	1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1		
Dd	418	KD-NLQVIEGOSKGLALELELDJLSN-QVHILPRKLPQGLKGLYLLISNRKLALPRA	475				
QY	351	-----HQQVYESLPN	-----	LDQVGLHG	366		
		1	1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1		
Dd	476	DALGLOARWDSVSHNRLEALPNSLAPLGRKRTSLRNNSLRTPTPQPGJERKMLEG	533				
QY	369	NPTRCDVIR-----WANATGRV-RETEP	-----	OSTCAEPDLOALFVR	409		
		1	1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1		
Dd	536	NPWDDGCFKLKLDLDFALONPSAVREFVDAICBDDCCOPATYTNNTICASBPVVGDLR	595				
QY	410	EV	411				
		1	1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1		
Dd	596	DL	597				

RESULT 6
US-08-473-089-49
Sequence 49, Application US/08473089
Patent No. 6342368
GENERAL INFORMATION:
APPLICANT: Mochly-Rosen, Daria
APPLICANT: Ron, Dorit
TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
TITLE OF INVENTION: Theeef
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 2000 Pennsylvania Avenue, NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/473,089
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 2550-0025.22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 605 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: Insulin-like growth factor binding
INDIVIDUAL ISOLATE: protein complex, Fig. 32
US-08-473-089-49

Query Match 10.9%; Score 407.5; DB 4; Length 605;
Best Local Similarity 25.2%; Pred. No. 6,1e-29;
Matches 152; Conservative 72; Mismatches 173; Indels 205; Gaps 16;
QY 3 LVAPELLAMVA-----GATATVTVVPMHVCPPQCACQIRPWTTPRSSY---REATT 52
DB 8 LALALLLSVALGPRSLGADPGTGEAGPCAPACVC-----SYDDADELS 57
QY 53 VDCNDLFTAVPPALPAGTQTLLQNSIVRV-----DOSEIGY----- 91
DB 58 VFCSRRUTLPLDGVPGGTALWLDGNNLSVPPAPQNISSIGFTLNLGGQGLSLEPQA 117
QY 92 ---LANTELDELSONSFSDARDDCFHALPOLSLHLEENOTRLE----- 133
DB 118 LIGLENCCHLERNQNRSLALGTFAHPALASLGSNNLSRLDGLFPGLSLWDLN 177
QY 134 -----DHSFAGLASLOELVNLNOLYRIAPRAFSGSLNLRHLNLSNLLRAIDSR 184
DB 178 GWSLAVLPAPAFRGSLSELVLAGNRLAYLDPALEFSGLAELRELDLSNALLRAIKANY 237
QY 185 FEMLPNLEIIMIGGNKVDAT-----LDNN----- 208
DB 238 FVQPRLOKXLDNRNLIAAFAFGFLKALNRWLDLSHNNAVAGLDETFPGGLGLRYRL 297
QY 209 -----FRPLANLSRLVLAGNMLREISDYALEGLQSLSSTFYDNLARY---- 252

DB 298 SHNAIASLRPRTRFDHLELEQIGHNRIRQLAERSPEGLGLEVLTLHNOLOEVKAGA 357
QY 253 -----PRALAEVPGKTLFLDKNPLOLVGCGDFANMLHKELEIG 292
DB 358 FLGLTVNAVAMNLSGNCRLNPLBOVFRGLKLSHLESGSLGRIRPTFTGLSLRRL 417
QY 293 NMDELVSIDKFAVLNPELTIKLDTNNPRLSFTHPRAFHNLPMQETIMLNNAALSAL-- 350
DB 418 KD-NGLVIERQSLMGSLAELELDITSN-QTLPHRLPGCLGKLEVLSSRRRLAEPLA 475
QY 351 -----HQQTVESLPN-----LQEVGLHG 368
DB 476 DALGPLQRAFWLDVSHNRLEALPNSLAPLGRUYLSLRNNSLRTFTPPQGLERLMEG 535
QY 369 NPIRCDCVIR-----WANATGRV--RTEP-----QSTLCAPPDQLRLPVR 409
DB 536 NPWDGCGPLKALRDFALQNSAVPRFVQAICEGDDCOPPAYTNNITCASPEVYGLDLR 555
QY 410 EV 411
DB 596 DL 597
RESULT 7
US-09-063-950-2
Sequence 2, Application US/09063950C
Patent No. 6225085
GENERAL INFORMATION:
APPLICANT: Holzman, Douglas A.
TITLE OF INVENTION: NOVEL LRSQ PROTEIN AND NUCLEIC ACID MOLECULES AND USES
FILE REFERENCE: MEI-019
CURRENT APPLICATION NUMBER: US/09/063,950C
CURRENT FILING DATE: 1998-04-21
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 2
LENGTH: 673
TYPE: PRT
ORGANISM: Homo sapiens
US-09-063-950-2
Query Match 10.8%; Score 402; DB 4; Length 673;
Best Local Similarity 24.9%; Pred. No. 2,3e-28;
Matches 194; Conservative 73; Mismatches 306; Indels 206; Gaps 24;
QY 4 LVAPLLAMVAGATATVTVVPMHVCPPQCACQIRPWTTPRSSYREATTVDNCNLFLLAV 63
DB 7 LLLPLLLLAUG-----PGVQGCPSGCOS-----QPQTVFCTAROGTTV 46
QY 64 PPALPAGTQTLLQNSIVRVDOSELYLANLTELDELSONSFSDARDDCFHALPOLSLH 123
DB 47 PDDVPTDVGTLVFENGITMDAGSFAGLPGIQLDLSQNDIASLPSVFPPLANLSLD 106
QY 124 LEENOTRLEDSHFSAGLASLOELVNLNOLYRIAPRAFSGSLNLRHLNLSNLLRAIDSR 183
DB 107 LLANLHETNTETFRGLRLERLTYGKNRIRHIOGARDTDRLELKLONEDELRALPPL 166
QY 184 WPEMLPNLEIIMIGGNKVDAILDMANFRPLANLSVLVLAGMMLRELSYALGLQSLSELS 243
DB 167 ---RLPRLILDLSHNSSLA--LEPGILDTAVNEALRAGLGLQIQDEGLFSRLRLHLD 222
QY 244 FYDNCOLARYRRALQVYGLKFLDKNPLOLVGPGDFANMLHKELELNMEELVSIDK 303
DB 223 VSDNOLERYP-----PVIR----- 236
QY 304 FALVNLPELTIKLDTNNPRLSFTHPRAFHNLPMQETIMLNNAALSALHQTVESLPNIOE 363
DB 237 ---GLRGTLRLRAGNRIARLQRPEDLAGLAALDELVSNSLSQALPFGDSLGLFRLRL 292
QY 364 VGLHGMPITGDCVIRMAN--ATGTRVRFIEPQSTLCAPPDQLRLPVRVEVFRREM----- 416

```
Db 293 LAARPFNCVCPILSMFGWRESHYTLASPEETRCFFPKNAGRLLLELDYADFQCPAT 352
QY 417 -TQHCPLISPRSPFSLQVASESMVILHORALEPEPELYWTWPAGLLTP----- 467
Db 353 TTTATVPTTTPVYREPTALSSSLAPTWLSFTATAPSPPTAPPTVGPVPODCCPS 412
QY 468 -----AAGRRYR---VYPEGTLELRVYTAEEAGLYTCVQNLVGAQDTKTVSVVGRA 517
Db 413 TCLNGCTCHGTGRHNLACLPCEGF-----TGLY-CEGQ----- 444
QY 518 LLDGPGDEGGELRYOEHPHPIILSWTPPTVSTNLTWSSASLRG-----QCAT- 570
Db 445 -----MGQTRSPPTVTP-----RPPRSLLTGLEPVSPTSLRVGLQRYLQSSSV 489
QY 571 -----ALALPRGTHSYNTRRLQATEYMACL-----OYA 600
Db 490 QLRSLRLTYRNLSGPRKRLVTLRLPASLAETVYTLQRPNATYVCVMPGPRVPEGEEA 549
QY 601 FADAHTQLACWARTKEATSCHRALGDRPGLIATLAVLLAAGLAHLGTGPRKGVG 660
Db 550 CGEATTPPA---VHSNHAIVTQAREGNLPLLIA-PALAAVLLAA-LAAVGAAYCVBRG-- 602
QY 661 GRPRLPAAWFGWASPSRVVSAPLYL-----PNNPGRK-----LPRSSSEETLL 706
Db 603 --RAAAAAADKQGVGP-----AGPLELEGVAVPLEPGKATEGGEGALPSGECEVPL 655

RESULT 8
US-09-063-950-5
; Sequence 5, Application us/09063950C
; Patent No. 6225085
; GENERAL INFORMATION:
; APPLICANT: Holzman, Douglas A.
; TITLE OF INVENTION: NOVEL LRGG PROTEIN AND NUCLEIC ACID MOLECULES AND USES
; FILE REFERENCE: MEI-019
; CURRENT APPLICATION NUMBER: US/09/063,950C
; CURRENT FILING DATE: 1998-04-21
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 605
; TYPE: PRT
; ORGANISM: Papio hamadryas
US-09-063-950-5

Query Match 10.7%; Score 397.5; DB 4; Length 605;
Best Local Similarity 24.9%; Pred. No. 5,1e-28;
Matches 150; Conservative 73; Mismatches 173; Indels 207; Gaps 16;
```

```
QY 3 LLAVALLLAAVNA-----GATATVPVYVPMWPCPPQACQIRWYTPRSSYREAT--T 52
Db 8 LALALLLSWVALGPRSLGAEPTGPEAGEPACPATCAC-----SYDDEVNELS 57
QY 53 VVCNDLFLTAVPALPAGTQTLLOSNTIVY-----DQS 87
Db 58 VFCSSRNLTLPDGIPTGQALWLDSSNNLSSIPPAFRNLSSLAFLNLQGGQLSLEPQA 117
QY 88 ELGYLANLTLELDLSONSFSDARDCDFHALPOLLSLHEENQITRL----- 133
Db 118 LLG-LENICHLERNQRLSLAVGFATTPALALLIGLSNNLSRLSDGLFEGGLNMDLN 176
QY 134 -----DHSRAGLASLOELVYNHQLYRIAPRAFSGLSNLRLHLNSNLRAIDSR 183
Db 177 LGWNSLAVLPAAAFRGLGELRELVLAGNRLAYLQPALFSGLAEIRELDLSNNALRAIKAN 236
QY 184 WPEMPLNLEIIMIGNKVDAI-----LDNM----- 208
Db 237 VFAQLRLQKLYLDRNLIAAVALPAGFGLKALRWLDLSHNVAIGLEDTPEPGLGLRVL 296
QY 209 -----FRPLANLRSVLAGMNLREISDYALEGLSLESISFYDNQALARY--- 252
```

```
Db 297 LSHNIAIARPRTEBDLHFLLELQLGHNRRITQLARESPFEGIGQLEVTLDHNOLOQEVK 356
QY 293 -----PRALQVPGKLFKLDINKNPLQVGVGDPAANMLHLKEIG 291
Db 357 AFLGLTVNAVNNLSONCLRNLPQGVFRGLKHLHLEGSCLGRIRBHTFAGLSGLERLF 416
QY 292 LNNHEIVSIDKFLVNLPELTKIDITNNPRLSTIHRRAHHLPQMTMLNNALSAI- 350
Db 417 LKD-NGVLGEQSLWGLAELELDLTSN-QLTHLPOLFOGLGKLEYLLLSHRLAELP 474
QY 351 -----HOQVEESL-----PILQEVGLH 367
Db 475 ADALGPLQRAFWLDVSHNRLEALPGLSLASLGRRLYLNLRNNSLRITTPQPGELRLMLE 534
QY 368 GNPRLRCQVIR---WANAGTRY-RLEP-----QSTLCAEPDQLRLPY 408
Db 535 GNPWDCSCPLKALRDFAIOWPSAVPRFVOAICEGDDCOPVYTYNNITCASPEVAGIDL 594
QY 409 REV 411
Db 595 RDL 597

RESULT 9
US-09-182-024A-2
; Sequence 2, Application us/09182024A
; Patent No. 6342370
; GENERAL INFORMATION:
; APPLICANT: Connolly, Timothy
; APPLICANT: Rajput, Bhanu
; TITLE OF INVENTION: Human S11t Polypeptide and Polynucleotides Encoding
; FILE REFERENCE: 640100-271
; CURRENT APPLICATION NUMBER: US/09/182,024A
; CURRENT FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: 60/063,946
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/096,420
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1523
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-182-024A-2

Query Match 9.5%; Score 355.5; DB 4; Length 1523;
Best Local Similarity 21.8%; Pred. No. 1,8e-23;
Matches 154; Conservative 94; Mismatches 259; Indels 199; Gaps 18;
```

```
QY 13 VAGTATVPVYVPMWPCPPQACQIRWYTPRSSYREATTYVQNDLFLTAVPALPAGIQ 72
Db 18 LALALASVLSGPPAAVACPCTKC-----SAASVQDHGGLRAVPRGIPRAAE 64
QY 73 TLLQNSIYRVDOSEIGYLANLTLELDLSONSFSDARDCDFHALPOLLSLHEENQITRL 132
Db 65 RLDDLNRNITRITRKDPFAGIKNRLVHLIEDNOYSVIERAGFODLQLETLRLNKKLQYL 124
QY 133 EDHSFAGLASLOELVYNHQLYRIAPRAFSGLSNLRLHLNSNLRAIDSRMEPLNLE 192
Db 125 PELIFQSTRLTRLDLSENOIGIPRKARGITDVYKNIQLDNNHISCIEDGAFALRDLE 184
QY 193 ILMIGNKVDAIILDMNFRLANLRSI----- 218
Db 185 ILTLNANNNISRIILVISFNHMPKIRTLRLSHNLYCDCHLAWLSDWLROKRVYQGFQLCMA 244
QY 219 --VLAGMNLREIS-----DYALEGL----- 236
Db 245 PVHLRGFNADVQKREYVCPAPRHSPPSCNANSISCPSPCTCSNNIIVDCRGKGLMEIFAN 304
```

```

US-08-190-802A-50

Query Match          9.2%: Score 344; DB 1; Length 603;
Best Local Similarity 23.3%: Pred. No. 4,56-23;
Matches 141; Conservative 63; Mismatches 194; Indels 208; Gaps 16;

OY 8 LLLAWA-----GATATVPVPMHVCPPQCAQCQIRPMYTPTRSSRYREATVYDCN 56
   |||||
Db 13 LLAWVALGRCGLQSTDPGASADAE---GQCPVACTCS-----HDDITBELSVFCS 61
   |||||

OY 57 DLFTAPAPALPAGTOTLLQSNISIVR-----DQSEBGY 91
   |||||
Db 62 SKNLTFLPDDIPIVSTALMDGNNISIPSAFQNLMSLDFLNLQSGWLSRSLERQALIG- 120
   |||||

OY 92 LANITLEDLSQNSPDRDCDFHALPOLISLLENNOTRLE----- 133
   |||||
Db 121 LQNTLYLHEERNRLRMVAVGLFHTTPSLASLSLSSNLGRLEBGLFQGLSHLMDNLGNM 180
   |||||

OY 134 -----DHSEVAGLASLOEYLNHNNOLYRIAPAFSGLSNLRHLHNSNLRAIDSRMFEM 187
   |||||
Db 181 SLVLPPTVQGLGNLHELVLGAKKITLYQPALFQGLQELAEIDLSSNALRSKANYFVN 240
   |||||

OY 188 LPNLELIMIGNKVDAI-----IDMN----- 208
   |||||
Db 241 LPRLQKLYLDRLNLITAVAPGAFLGKMKALRWLDLSHNRVAGIMEDTPGLGLHVLRLAHN 300
   |||||

OY 209 -----PRPLANRSLVLAGMNLRELSDALGLOSLESFYNOLARV----- 252
   |||||
Db 301 AIASLRPRTERKDLHLEELQGLGHNRRKIQGERTFEGQLQLEVLTLNDNQITREVRVCAFGS 360
   |||||

OY 253 -----PRALAEQVPGLEKFLDNLNKPQVRPGDFANMLHRELGLNMM 295
   |||||
Db 361 LFNVAVMNLGNCRLSRIPERVFGDLKLSHLSEHSCIGHVRLHTFAGLSGLRRLFLRD- 419
   |||||

OY 296 EELVSIDKFAVLNLPETIKDITNNRSLFTIPRAFHNLPMOETLMLNNRSL- 350
   |||||
Db 420 NMSISIEOSLAGISELLELDITTN-RLTHLPRQLFQGLGHEVLLISYQNLTTLSAEVL 478
   |||||

OY 351 -----HQ-----QTVESLPNLQEVGLHGNPIR 372
   |||||
Db 479 GPLQRAWLDLSNNHLETLAEGFLSSLGVRVYLSLRNNSLDTFPSQGLBETMLDANPMD 538
   |||||

OY 373 CDCVIR-----WANAGTRVRFTEP-----QSTLCAPPPDLQRLPVREVPFR 414
   |||||
Db 539 CSCPLKALRDFALQNGVVRVQTVCEGDQCPYYTYNNITTCGAPANVSGLDLADVSET 598
   |||||

OY 415 EMTDHC 420
   ||
Db 599 HFV-HC 603

RESULT 11
US-08-477-346-50
; Sequence 50, Application US/08477346
; Patent No. 6262023
; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; TITLE OF INVENTION: Theeof
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morison & Foerster
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```



```

Db      241 LPRLOKLYLDRNLITLVAAPGAFGLMKRLRWLIDLSHNKVAIGMEDTFEGLLGLHVLRLAHN 300
QY      209 -----PRPLANLRSLVLAGNNLRREISDYALMEGLQSEISLSPFNOLARY----- 252
Db      301 AIASLRPRTEKDFHLFEELQDGHNRITQDLERFFEGSGOLEVLTLDNQITVEYRVGAFSG 360
QY      253 -----PRRALEQVPGLKFLDLINKNPLQGVGSGDEAFNNMLHLKELGINNN 295
Db      361 LFNVAVMNLSCNCLRSIPERFEGQDLKHSIHLESGGLGHVRLHRTFGSLGSLRRFLRD- 419
QY      296 EELVSIDKFAVLNLRPELTLDITNNPLRSLTHPRAFHHLRPMETLMINNALSAL----- 350
Db      420 NSISIEEOSLAGSELLELDTTN-RLTHPLRQLFQGLHLEVLITLSTYQDLTSLAEVL 478
QY      351 -----HQ-----QTVESPLNLEQVGHGMPIR 372
Db      479 GPLQRAFWLDSHHHLETLAEGLFSSIGRVRYLSLRNNSIQDTSSPOGLERLWIDANPMD 538
QY      373 CDQYIR-----WANATGTRVAFLEIP-----QSTLCAEPDLQRLPVRREVPR 414
Db      539 CSCPLKALRDFALQNPGVVPRFVQVTOEGDDCQPVYTYNNITCGAPANVSGLDLRDVSE 598
QY      415 EMTDHC 420
Db      599 HFV-HC 603

```

```

RESULT 13
; Sequence 2, Application US/09191647
; Patent No. 6046015
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/191,647
; CURRENT FILING DATE: 1998-11-13
; EARLIER APPLICATION NUMBER: 60/065,544
; EARLIER FILING DATE: 1997-11-14
; EARLIER APPLICATION NUMBER: 60/081,057
; EARLIER FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1525
; TYPE: PRT
; ORGANISM: human
; US-09-191-647-2

```

Query Match	8.9%;	Score 331;	DB 3;	Length 1525;
Best Local Similarity	17.8%;	Pred. No. 3.4e-21;		
Matches 166; Conservative	76;	Mismatches 177;	Indels 330;	Gaps 11;

[illegible]

```

QY 205 -----LDNM 208
      | : |
Db 255 FVCSDEEGHOSFMAPSCSVLHCPAOTCSNNIYDCRGKLTPTNLPETITEIRLEON 314
QY 209 -----FRPLANIRSLVLAGNNLEISDYALEGOSLESSEFYNOIARVPRR----- 255
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 315 TIKYIPGEAFSPYKKLRKRIDLSNNOISBELADAFQGRLSNLSVLYXNKTKTELPKSLFEG 374
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 256 -----ALEQVPGIKELDLNKNPLQVRGQDFEAM----- 284
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 375 LFSQLLLNANKINCLEVDFAFODLHNILNLSYDNKLTQIAGTFSPRLAIOTMHLAON 434
QY 285 -----LHK----- 288
      | |
Db 435 PFICDCHLKMLADVLHTNPITETSGARCTSPRLANKRIQOTSKSKFFCSGTEGYRSKLSG 494
QY 289 -----ELGANNMEEIWSIDRFALYNLP 310
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 495 DCFADLACPEKRCRGCTVDCSNOKLNKIREPHIROYTAELNENNEPTVEATGIEFKPL 554
QY 311 ELTKLDTITNNP-----RLSIRHRAFNHLPOWETLMLNNAL 347
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 555 QIRKINFNNKNTIIEGAFEGASGVNEILLTSNMLEVNOHKMKRGLSLEITLMSNRI 6144
QY 348 SALHQOT-----VESLPNLOEVLGHNPIRCDVYIRMANAT 383
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 615 TCVGNDSFGLGSVRLLSYDNQITTVAPGAFDTLHSLTLLNLANPENCYIAMLGEW 6744
QY 384 GTRVRFI--EQOSTICARPPDQRLRPVAREVFRRENT-----DHCIPL 423
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 675 LRRKRIIVGNPR--CQKPYFLKEIPIODVAIADPTCDGCDNDNSCSFL 720

```

```

1  RESULT 14
2  / Sequence 2, Application US/09540245A
3  / Patent No. 6270984
4  / GENERAL INFORMATION:
5  / APPLICANT: Goodman, Corey
6  / APPLICANT: Kid, Thomas
7  / APPLICANT: Brose, Katja
8  / APPLICANT: Tessier-Lavigne, Marc
9  / TITLE OF INVENTION: Modulating Robo: Ligand Interactions
10 / FILE REFERENCE: B98-031-3
11 / CURRENT APPLICATION NUMBER: US/09/540.245A
12 / CURRENT FILING DATE: 2000-03-31
13 / PRIOR APPLICATION NUMBER: 60/065,544
14 / PRIOR FILING DATE: 1997-11-14
15 / PRIOR APPLICATION NUMBER: 60/081,057
16 / PRIOR FILING DATE: 1998-04-07
17 / NUMBER OF SEQ ID NOS: 20
18 / SOFTWARE: PatentIn Ver. 2.0
19 / SEQ ID NO: 2
20 / LENGTH: 1525
21 / TYPE: prt
22 / ORGANISM: human
23 / US-09-540-245A-2

```

Query Match	8.9%	Score 331	DB 4	length 1525
Best Local Similarity	17.8%	Pred. No. 3.4e-21		
Matches 126; Conservative	76;	Mismatches 177;	Indels 330;	Gaps 11

```

QY      29  CPGGACQGLRPWYTPRSTSSYREKATVYDNDCLFLTAAPPAQTCGLLLNSNIYVDSDS  88
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      28  CPGQCC-----SSGYDCHGLALRSYPRNPRTRELDLNGNITRTYKTD  74
QY      89  LGLTANLTELDSQNSFSPARDCDFHALPQLLSLHLEQQLTRLDHSPAGIASIQEYL  144
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      75  FAGIRLRLRYQLEMENITSTERGAFODLEKELERLRNNRHLDLPPELLFLGTAKYLRDL  133
QY      149  NHQQLYRIAPRAFSGLSNLRLLNSLRLLRAIDSRKEFEMLPLELTITMGKNVDAI----  203

```

```
Db 135 SENOIOAIRKRAFGAVDIKNIQLDYNOJSCIEDGAFRALRDLVLTNNNNTRLSVAS 194
QY 205 ----- 204
Db 195 FNNHMKLRTFRLHSNNLYCCHLAWLSDMLKRRPRVGLYTQCMGSPSHAGHNAEYOKRE 254
QY 205 -----LDNN 208
Db 255 FVCSDEEGHOSFMAPSCSVLHCPAAGTCSNNIVDCRGKGLTEIPTNLPETITEIRLEON 314
QY 209 -----FRPLANLSVLAGMNLREISDYALBGLQSLSESIFYDNOIARVPRR---- 255
Db 315 TIKVITPGAFSPYKRLRRIIDLSNNQISELAPDAFQGLRSINSIVLYGNKITELPKSLFEG 374
QY 256 -----ALEOYPLKFLDLINKNPLOVFGPDFAAN----- 284
Db 375 LFSIQLLLNANKINCLRVDAFQDLHNLNLSYDNKLQTIKGTSPRLAIQTMHLAON 434
QY 285 -----LHLK----- 288
Db 435 PFICDCHLKWADLYLHNPITETSGARCTSPRLANKRIGQIKSKKFCSTEDYRSKLSG 494
QY 289 -----ELGNNMEELYSIDKFAVNI 310
Db 495 DCFADLACEPKRCCEGTVDCSNOKLNKIPBEHPOYTAEURLNNEFTVLEATGIFPKLP 554
QY 311 ELTKLIDITNPN-----RLSFTHPRAFHILPOMETLMLNNAL 347
Db 555 QLRKINSNNKITDIEGAFEGASGVNEILLTSNRLENVOHKMFKGLESKTLMLRSNRI 614
QY 348 SALHQOT-----VESLPNLOEVGLHGNPIRCDCVIRMANAT 383
Db 615 TCVGNDSFGLSSVRLSLYDNOITTVAPGAFDTLSLSTLMLANPENCYLAWLGM 674
QY 384 GTRVRFI--EPOSTLCAEPDLORLPYREVPRFEMT-----DHCLPL 423
Db 675 LRKRRIYGNPR--CQKPYFLKEIPIDVAIODEFTCDGNDNCSPL 720

RESULT 15
US-09-540-153-2
; Sequence 2, Application US/09540153
; Patent No. 6270995
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/540,153
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/191,647
; PRIOR FILING DATE: 1998-11-13
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1525
; TYPE: PRT
; ORGANISM: human
US-09-540-153-2
```

Query Match 8.9%; Score 331; DB 4; Length 1525;
Best Local Similarity 17.8%; Pred. No. 3,4e-21;
Matches 126; Conservative 76; Mismatches 177; Indels 330; Gaps 11;

```
QY 29 CPPOCACQIRPMWYPRSSYREATTVCNDLFLTAVPPALPAGTQTLQLQSNISIVRQSE 88
Db 28 CPAQSCS-----SGSTVDCGHALRSVYRNIPRNTERTLDLNGNNITRIITKTD 74
```

```
QY 89 LGYLANLTFELDSQNSFSDARDODFHALPOLLSLHLENOLTRLEDHSPAGIASLOETYL 148
Db 75 FAGLRLAVLOLMEKKISTIRGAFODLKEERLRNLNHLQLEPPELLFGATAKLYL 134
QY 149 NNNOLYRIAPAFSGSLNLRHLNSNLLRAIDSRWEMLPNLEIIMIGSKYDAI----- 204
Db 135 SENOIOAIRKRAFGAVDIKNIQLDYNOJSCIEDGAFRALRDLVLTNNNNTRLSVAS 194
QY 205 ----- 204
Db 195 FNNHMKLRTFRLHSNNLYCCHLAWLSDMLKRRPRVGLYTQCMGSPSHLGHNAEYOKRE 254
QY 205 -----LDNN 208
Db 255 FVCSDEEGHOSFMAPSCSVLHCPAAGTCSNNIVDCRGKGLTEIPTNLPETITEIRLEON 314
QY 209 -----FRPLANLSVLAGMNLREISDYALEGLQSLSESIFYDNOIARVPRR---- 255
Db 315 TIKVITPGAFSPYKRLRRIIDLSNNQISELAPDAFQGLRSINSIVLYGNKITELPKSLFEG 374
QY 256 -----ALEOYPLKFLDLINKNPLOVFGPDFAAN----- 284
Db 375 LFSIQLLLNANKINCLRVDAFQDLHNLNLSYDNKLQTIKGTSPRLAIQTMHLAON 434
QY 285 -----LHLK----- 288
Db 435 PFICDCHLKWADLYLHNPITETSGARCTSPRLANKRIGQIKSKKFCSTEDYRSKLSG 494
QY 289 -----ELGNNMEELYSIDKFAVNI 310
Db 495 DCFADLACEPKRCCEGTVDCSNOKLNKIPBEHPOYTAEURLNNEFTVLEATGIFPKLP 554
QY 311 ELTKLIDITNPN-----RLSFTHPRAFHILPOMETLMLNNAL 347
Db 555 QLRKINSNNKITDIEGAFEGASGVNEILLTSNRLENVOHKMFKGLESKTLMLRSNRI 614
QY 348 SALHQOT-----VESLPNLOEVGLHGNPIRCDCVIRMANAT 383
Db 615 TCVGNDSFGLSSVRLSLYDNOITTVAPGAFDTLSLSTLMLANPENCYLAWLGM 674
QY 384 GTRVRFI--EPOSTLCAEPDLORLPYREVPRFEMT-----DHCLPL 423
Db 675 LRKRRIYGNPR--CQKPYFLKEIPIDVAIODEFTCDGNDNCSPL 720
```

Search completed: August 31, 2002, 14:35:27
Job time: 29 sec

Sat Aug 21 14:55:21 2002

us-09-905-088a-245.ra1

Page 11

PR 18-SEP-1997; 970S-0059263.
PR 18-SEP-1997; 970S-0059266.
PR 15-OCT-1997; 970S-0062125.
PR 17-OCT-1997; 970S-0062285.
PR 17-OCT-1997; 970S-0062287.
PR 21-OCT-1997; 970S-0063486.
PR 24-OCT-1997; 970S-0062814.
PR 24-OCT-1997; 970S-0062816.
PR 24-OCT-1997; 970S-0063045.
PR 24-OCT-1997; 970S-0063120.
PR 24-OCT-1997; 970S-0063121.
PR 24-OCT-1997; 970S-0063127.
PR 24-OCT-1997; 970S-0063128.
PR 27-OCT-1997; 970S-0063329.
PR 27-OCT-1997; 970S-0063327.
PR 28-OCT-1997; 970S-0063541.
PR 28-OCT-1997; 970S-0063542.
PR 28-OCT-1997; 970S-0063544.
PR 28-OCT-1997; 970S-0063549.
PR 28-OCT-1997; 970S-0063550.
PR 28-OCT-1997; 970S-0063564.
PR 29-OCT-1997; 970S-0063435.
PR 29-OCT-1997; 970S-0063704.
PR 29-OCT-1997; 970S-0063732.
PR 29-OCT-1997; 970S-0063738.
PR 29-OCT-1997; 970S-0063734.
PR 29-OCT-1997; 970S-0064215.
PR 29-OCT-1997; 970S-0063735.
PR 31-OCT-1997; 970S-0063870.
PR 31-OCT-1997; 970S-0064103.
PR 03-NOV-1997; 970S-0064248.
PR 07-NOV-1997; 970S-0064809.
PR 12-NOV-1997; 970S-0065186.
PR 17-NOV-1997; 970S-0065846.
PR 18-NOV-1997; 970S-0065693.
PR 21-NOV-1997; 970S-0066120.
PR 21-NOV-1997; 970S-0066364.
PR 24-NOV-1997; 970S-0066772.
PR 24-NOV-1997; 970S-0066466.
PR 24-NOV-1997; 970S-0066770.
PR 24-NOV-1997; 970S-0066511.
PR 24-NOV-1997; 970S-0066453.
XX
PA (GENE) GENENTECH INC.
XX
PI Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;
XX
DR MPI: 1999-229533/19.
XX N-PSDB; AAX52256.
PT New isolated human genes and polypeptides used in, e.g. treatment of
XX gastrointestinal ulceration
PS
XX Claim 12; Fig 86; 320pp; English.
XX
CC AAY13344-403 represent secreted and transmembrane human proteins.
CC The cDNA sequences are obtained from cDNA libraries, prepared from
CC fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
CC The encoded polypeptides have specific uses based on their homology to
CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
CC associated with the preservation and maintenance of gastrointestinal
CC mucosa and the repair of acute and chronic mucosal lesions
CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
CC ulceration and congenital microvillus atrophy), skin diseases associated
CC with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial
CC cancers such as lung squamous cell carcinoma of the vulva and gliomas),
CC potent effects on cell growth and development, diseases related to
CC growth or survival of nerve cells including Parkinson's disease,
CC Alzheimer's disease, ALS, neuropathies or cancer. PRO265 can be used as
CC a target for anti-tumor drugs. PRO53 may be used in the treatment
CC of Usher Syndrome or Atrophia areata. PRO269 can be used as an
CC anti-thrombotic agent; PRO287 polypeptides and portions may have

CC therapeutic applications in wound healing and tissue repair; PRO317 can
CC be used for treating problems of the kidney, uterus, endometrium, blood
CC vessels, or related tissue, e.g. in the heart of genital tract.
XX
SQ Sequence 713 AA:
Query Match 100.0%; Score 3732; DB 20; Length 713;
Best Local Similarity 100.0%; Pred. No. 1e-290;
Matches 713; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MRLVAPLLAMVAGATATVVPVPMHVCPPQCAQIRPWTYPRSSYREATVDCNDLFL 60
DB 1 mrlvapl llawagata t vvpvpmhvcppqccqirpytpssyre atvdcndlfl 60
QY 61 TAVPALPAGTQTLLILOSNSIVRVDSGLGYLANLTLFELDSQNSFSPARCDDHALPQL 120
DB 61 t avpalp agtqtll l osnsi vrvds gl ylanl t l f eldsqns f s parcd dhalpql 120
QY 121 SLHLEENQTRLEDSHFAGLASLOELYLNHNOLYRIAPRAFSGISNLRHLNSNLRAT 180
DB 121 sl hle enq trl eds h f aglas lo el yln h no ly ri a p raf sg is nlr h l ns nlr at 180
QY 121 sl hle enq trl eds h f aglas i gely lnnq ly ri a p raf sg is nlr i h l ns nlr at 180
QY 181 DSRWFEMLPNLEIMIGNKVDAILDMNRPRLANRSLVLAGMNLREISYALEGLQSL 240
DB 181 ds rwf em l p nle i m i g n k v d a i l d m n r p r l a n r s l v l a g m n l r e i s y a l e g l q s l e 240
QY 241 SLSPFYDNLAVRRAL EOYVGLKFLDLNKNP LQORVPGPFAMNLHKEGLNNMETVS 300
DB 241 sl spfy d n l a v r r a l e o y v g l k f l d l n k n p l q o r v p g p f a m n l h k e g l n n m e t v s 300
QY 241 sl spfy d n l a v p r r a l e q v p g l k f l d l n k n p l q r v p g p f a m n l h k e g l n n m e e t v s 300
QY 301 IDKFLVNLPELTKLDITNNPRLSFIRPAFHILPQMETLMLNNNLSALHQOTVESLPN 360
DB 301 i d k f a l v n l p e l t k l d i t n n p r l s f i r p a f h i l p q m e t l m l n n n l s a l h q o t v e s l p n 360
QY 361 LQEVGHGNPFRCCVIRMANATGTRVFTERPOSTLCAEPDQRLPVRVPRFRMTDHC 420
DB 361 l q e v g h g n p f r c c v i r m a n a t g t r v f t e r p o s t l c a e p d q r l p v r v p r f r m t d h c 420
QY 421 LPLSPRSFPSTLOVAGSESVNLGRALAEPEETVWTFAGRLTPAHAGRRYRVPEG 480
DB 421 l p l s p r s f p s t l o v a g s e s v n l g r a l a e p e e t v w t f a g r l t p a h a g r r y r v p e g 480
QY 481 TLELRVYAEAGLYTCVAQNLVGDATKTVSVVYGRALLQPGDESGGLELRYOETHPYH 540
DB 481 t l e l r v y a e a g l y t c v a q n l v g d a t k t v s v v y g r a l l q p g d e g g l e l r y o e t h p y h 540
QY 541 ILLSWTPPTNVTGNLWSSASSLRGGATATLRLPRGTHSYNITRLQATEWACLOVA 600
DB 541 i l l s w t p p t n v t g n l w s s a s s l r g g a t a t l r l p r g t h s y n i t r l q a t e w a c l o v a 600
QY 601 FADAAQLACVMARTEKATSCHRALGDRPLIALILAAVLLLAAGLAHIGTGQPRKGVG 660
DB 601 f a d a a q l a c v m a r t e k a t s c h r a l g d r p l i a l i l a a v l l l a a g l a h i g t g q p r k g v g 660
QY 661 GRRLPPAPAMAFWGSAPSVRVVSAPLVLPWNPGRKLP RRSSEGETLLPPLSQNS 713
DB 661 g r r l p p a p a m a f w g s a p s v r v v s a p l v l p w n p g r k l p r r s s e g e t l l p p l s q n s 713
RESULT 2
AAB80253 standard; Protein: 713 AA.
ID AAB80253
XX
AC AAB80253;
XX
DT 24-APR-2001 (first entry)
XX
DE Human PRO293 protein.
XX
KW Human; PRO: dermatological; antipsoriatic; cytostatic; antiinflammatory;
KW antiparkinsonian nootropic; neuroprotective; vulnary; cardiac;
KW antiangiogenic; vasotropic; antiasthmatic; antineumatic; cancer;

KM antiarthritic; antiinfertility; antidiabetic; antiviral; diabetes;
 KW ophthalmological; gene therapy; skin disease; gastrointestinal disorder;
 KM ischaemia; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN WO200104311-A1.
 XX
 PD 18-JAN-2001.
 XX
 PF 22-FEB-2000; 2000WO-US04414.
 XX
 PR 07-JUL-1999; 99US-0143048.
 PR 26-JUL-1999; 99US-0145698.
 PR 28-JUL-1999; 99US-0146222.
 PR 08-SEP-1999; 99WO-US20594.
 PR 13-SEP-1999; 99WO-US20944.
 PR 15-SEP-1999; 99WO-US21090.
 PR 15-SEP-1999; 99WO-US21547.
 PR 05-OCT-1999; 99WO-US23089.
 PR 29-NOV-1999; 99WO-US28214.
 PR 30-NOV-1999; 99WO-US28313.
 PR 16-DEC-1999; 99WO-US30095.
 PR 20-DEC-1999; 99WO-US30911.
 PR 20-DEC-1999; 99WO-US30999.
 PR 05-JAN-2000; 99WO-US00219.
 XX
 PA (GENE) GENENTECH INC.
 XX
 PI Ashkenazi AJ, Botstein D, Desnoyers L, Eaton DL, Ferrara N;
 PI Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;
 PI Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ, Kliaevin IJ;
 PI Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;
 PI Williams PM, Wood WI;
 XX
 DR WPI: 2001-081051/09.
 DR N-PSDB: AAF72414.
 XX
 PT Sixty one nucleic acids encoding PRO polypeptides which are useful in
 PT the treatment of skin diseases (e.g. psoriasis), cancers (e.g. lung
 PT squamous cell carcinoma) and neurodegenerative diseases (e.g.
 PT Alzheimer's disease) -
 PT
 PT
 XX
 PS Claim 1: Fig 86; 393pp; English.
 XX
 CC The present sequence is one of sixty one novel secreted and
 CC transmembrane PRO polypeptides. The PRO polypeptides are
 CC useful for treating skin diseases (e.g. psoriasis), cancers (e.g. lung
 CC squamous cell carcinoma), gastrointestinal disorders (e.g.
 CC enterocolitis), neurodegenerative diseases (e.g. Alzheimer's disease,
 CC Parkinson's disease), wound repair, cardiovascular disorders (e.g.
 CC endometrial bleeding angiogenesis, ischaemia such as coronary
 CC ischaemia, atherosclerosis), inflammatory disorders (e.g. asthma,
 CC rheumatoid arthritis, multiple sclerosis), infertility, AIDS and
 CC diabetes and retinal disorders such as retinitis pigmentosum.
 CC The PRO nucleic acids have applications in molecular biology, including
 CC use as hybridization probes, and in chromosome and gene mapping.
 CC
 XX
 SQ Sequence 713 AA:

Query Match 100.0%; Score 3732; DB 22; Length 713;
 Best Local Similarity 100.0%; Pred. No. 1e-290;
 Matches 713; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRLIYAPLLANVAGATITVPPVPHVPCPCQACQIPWPTTPSSSTREATTVCNDLFL 60
 DB 1 MRLIYAPLLANVAGATITVPPVPHVPCPCQACQIPWPTTPSSSTREATTVCNDLFL 60
 QY 61 TAVPPALPAGCTOTLLQSNSTVRVDSSEGLYANTTELDLSONSPSDARDDPHALPOL 120
 DB 61 TAVPPALPAGCTOTLLQSNSTVRVDSSEGLYANTTELDLSONSPSDARDDPHALPOL 120

QY 121 SLHLEENQTLRLDHSFAGLALQELLYLNHNOYLRIAPAFSGLSNLRHLNLSLLRAI 180
 DB 121 SLHLEENQTLRLDHSFAGLALQELLYLNHNOYLRIAPAFSGLSNLRHLNLSLLRAI 180
 QY 181 DSRWFEMLPNLETLMTGKNKVDAILDMNRPRLANLSVLVAGNNLEISDYALEGQSLE 240
 DB 181 DSRWFEMLPNLETLMTGKNKVDAILDMNRPRLANLSVLVAGNNLEISDYALEGQSLE 240
 QY 241 SLSEFYDQNLARVRRALEQPGKFLDLNKNPRLQRYPGDPFAMMLKELGNNMEELS 300
 DB 241 SLSEFYDQNLARVRRALEQPGKFLDLNKNPRLQRYPGDPFAMMLKELGNNMEELS 300
 QY 301 IDEKALVNPBELKLDITNNPRLSFIHPRAFHLPOMETLMNNLSALHQQTVESLPN 360
 DB 301 IDEKALVNPBELKLDITNNPRLSFIHPRAFHLPOMETLMNNLSALHQQTVESLPN 360
 QY 361 LQEVGLHGNPRICDCYIRMANATGTVRFRTEPOSTLCAEPDQLRLPREVPREKTDHC 420
 DB 361 LQEVGLHGNPRICDCYIRMANATGTVRFRTEPOSTLCAEPDQLRLPREVPREKTDHC 420
 QY 421 LPLISPRSPPSLOVASGESMVLHCRALAEPEEITVWTPAGRLPLPAHAGRRYRYPG 480
 DB 421 LPLISPRSPPSLOVASGESMVLHCRALAEPEEITVWTPAGRLPLPAHAGRRYRYPG 480
 QY 481 TLELRVYTAEEAGLYTCVAQNLGADTKTVSVVVGRRALLQPGHDESGQELRYOETHPYH 540
 DB 481 TLELRVYTAEEAGLYTCVAQNLGADTKTVSVVVGRRALLQPGHDESGQELRYOETHPYH 540
 QY 541 ILSWVTPPTVSTNLTWSSASSLRGOGATLALRPRGTHSVITRLLQNTERYMACLOYA 600
 DB 541 ILSWVTPPTVSTNLTWSSASSLRGOGATLALRPRGTHSVITRLLQNTERYMACLOYA 600
 QY 601 FADAHQOLACVMAKTKATKCHRALDPRGLTALLAVYLLAAGLAHNGTGPCKKGV 660
 DB 601 FADAHQOLACVMAKTKATKCHRALDPRGLTALLAVYLLAAGLAHNGTGPCKKGV 660
 QY 661 GRRLPPAMAFCWMSAPSVRVVSAPLVLPWNPGRKLPSSSEGTLLPPLSONS 713
 DB 661 GRRLPPAMAFCWMSAPSVRVVSAPLVLPWNPGRKLPSSSEGTLLPPLSONS 713

RESULT 3
 AAB92765 ID AAB92765 standard; Protein; 377 AA.
 XX
 AC AAB92765;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:11234.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI: 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT full-length cDNAs - of the abnormality of the proteins encoded by the
XX
XX Claim 8; SEQ ID 11234; 2537bp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 377 AA:
SQ

Query Match 53.4%; Score 1993; DB 22; Length 377;
Best Local Similarity 100.0%; Pred. No. 1.6e-151;
Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 337 METLMLNNLSALHOQTVESLPNLOEVGLHGNPIRCDVIRWANAATGTRVFTEPOSTL 396
DB 1 melimlnnalsalhgqtveslprlgevjhngpicdcvtrwamtgtrvrliepqstl 60
QY 397 CAEPPLQRLPVRVREPREKTDHCLPLISRSPSLQVANGSWMVLCRAIAEPEPEY 456
DB 61 caepplqrlprvrevpfremtdhclplisrpspslqvassgmvlhcralaapepeyl 120
QY 457 WTPPAGRLRPAHAGRRRYRYPREGTELRVTAEEAGITCVANONLVGADTKTVSVYVR 516
DB 121 wtppagrlrlpahagrrryrypegtelrrvtlaeeagilylcvagqlvgadtktvsvvygr 180
QY 517 ALLQPGDEGQGLRLVQETHRPYHILLSWTPPNTVSTNLTWSSASSLRGOGATALARLP 576
DB 181 allqpgdeggglelrvqethrpyhillswtppntvstnltwssasslrgogatalarlp 240
QY 577 RGTSHYNITRLLQATEWMAQLQVAFADAHTQACVARKTEATSCRRALGDRPGLTALA 636
DB 241 rgtshynitrlqatewmaqlqvafadahnqacvarkteatschrralgrdpgljalala 300
QY 637 LAVLLLAAGLAHLGTGQPRKGVGRRPLRPAMAFWGSNAPSRYVSAPLVLPWNGRKL 696
DB 301 lavlllaaglaahlgtgqprkgyvgrtrlpamafwgsnapsryvsaplvlpwngtrkl 360
QY 697 PRSSEGETLLPLPSONS 713
DB 361 prssegetllplpsqns 377

RESULT 4
AAB33472
ID AAB33472 standard: Protein; 716 AA.
XX
AC AAB33472;
XX

DT 29-JAN-2001 (first entry)
XX
DE Human PRO1338 protein UNQ693 SEQ ID NO:279.
XX
KW Human; immune related disease; diagnosis; antiinflammatory; cardiant;
KW dermatological; anhrathritic; antirheumatic; immunosuppressive;
KW haemostatic; antihypoid; antidiabetic; neutropilic; neuroprotective;
KW antinaemic; hepatotropic; vitucide; antiporiatic; antiallergic;
KW osteoarthritis; systemic lupus erythematosus; rheumatoid arthritis;
KW idiopathic inflammatory myopathy; Sjogren's syndrome; thyroiditis;
KW autoimmune thrombocytopaenia; autoimmune haemolytic anaemia; diabetes mellitus;
KW demyelinating disease; hepatobiliary disease; Whipple's disease;
KW inflammatory bowel disease; gluten-sensitive enteropathy;
KW autoimmune disease; immune-mediated skin disease; allergic disease;
KW immunological disease; transplantation associated disease;
KW graft rejection; graft-versus-host-disease.

Homo sapiens.

WO200053758-A2.

14-SEP-2000.

02-MAR-2000; 2000WO-US05841.

XX 08-MAR-1999; 99WO-US05028.
PR 10-MAR-1999; 99US-0123618.
PR 12-MAR-1999; 99US-0123957.
PR 23-MAR-1999; 99US-0125775.
PR 12-APR-1999; 99US-0128849.
PR 20-APR-1999; 99WO-US08615.
PR 28-APR-1999; 99US-0131445.
PR 04-MAY-1999; 99US-0132371.
PR 14-MAY-1999; 99US-0134287.
PR 02-JUN-1999; 99WO-US12252.
PR 23-JUN-1999; 99US-0141037.
PR 20-JUL-1999; 99US-0144758.
PR 26-JUL-1999; 99US-0145698.
PR 28-JUL-1999; 99US-0146222.
PR 01-SEP-1999; 99WO-US20111.
PR 08-SEP-1999; 99WO-US20594.
PR 13-SEP-1999; 99WO-US20944.
PR 15-SEP-1999; 99WO-US21090.
PR 15-SEP-1999; 99WO-US21547.
PR 05-OCT-1999; 99WO-US23089.
PR 29-OCT-1999; 99US-0162506.
PR 29-NOV-1999; 99WO-US28214.
PR 30-NOV-1999; 99WO-US28313.
PR 30-NOV-1999; 99WO-US28409.
PR 01-DEC-1999; 99WO-US28301.
PR 01-DEC-1999; 99WO-US28634.
PR 02-DEC-1999; 99WO-US28551.
PR 02-DEC-1999; 99WO-US28564.
PR 02-DEC-1999; 99WO-US28565.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30999.
PR 30-DEC-1999; 99WO-US31274.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00277.
PR 06-JAN-2000; 2000WO-US00376.
PR 11-FEB-2000; 2000WO-US03565.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.

(GETH) GENENTECH INC.

XX Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hebert C, Henzel W;
PI Kadakoff RC, Lu Y, Pan J, Pennica D, Shelton DL, Smith V;
PI Stewart JA, Tumas D, Watanabe CK, Wood WI, Yan M;
XX

DR MPI: 2000-572271/53.
DR N-PSDB: AAC58637.
XX
XX Sixty four PRO polypeptides, useful in the diagnosis and treatment of
PT immune related disorders, e.g. systemic lupus erythematosus, rheumatoid
PT arthritis, osteoarthritis, thyroiditis and diabetes mellitus -
XX
XX Claim 33: F1g 118: 309pp: English.
XX

The present invention describes sixty four human PRO proteins which can be used in the treatment of immune related diseases. The human PRO proteins, anti-PRO antibodies, agonists and antagonists are useful for treating and diagnosing immune related disorders. The disorders are selected from systemic lupus erythematosus, rheumatoid arthritis, osteoarthritis, juvenile chronic arthritis, spondyloarthritis, systemic sclerosis, idiopathic inflammatory myopathies, Sjogren's syndrome, systemic vasculitis, sarcoidosis, autoimmune haemolytic anaemia, autoimmune thrombocytopenia, thyroiditis, diabetes mellitus, immune-mediated renal disease, demyelinating diseases of the central and peripheral nervous systems, hepatobiliary diseases, inflammatory bowel disease, gluten-sensitive enteropathy and Whipple's disease, autoimmune or immune-mediated skin diseases, allergic diseases, immunological diseases of the lung, and transplantation associated diseases including graft rejection and graft-versus-host-disease.

AAC58397 to AAC58578 represent PCR primers and hybridisation probes used in the isolation of human PRO sequences. AAC58579 to AAC58642 and AAB33414 to AAB33477 represent human PRO polynucleotide and protein sequences given in the exemplification of the present invention.

[illegible]

Db 566 phitylavpvdvneynlthqpsbtqvevcltvsniinqgqkscvwnltknaafavaisd 6255

Qy 619 ----TSCHRALGDRPGLTIAITLAV 639

Db 626 qetstalaavmgsmfavistasiav 650

Result	Accession	Standard	Protein	GI
5	AA012427	standard	Protein	716 AA
AC	AA012427			
AD	AA012427			
AE	AA012427			
AF	AA012427			
AG	AA012427			
AH	AA012427			
AI	AA012427			
AJ	AA012427			
AK	AA012427			
AL	AA012427			
AM	AA012427			
AN	AA012427			
AO	AA012427			
AP	AA012427			
AQ	AA012427			
AR	AA012427			
AS	AA012427			
AT	AA012427			
AW	AA012427			
AX	AA012427			
AY	AA012427			
AZ	AA012427			
BA	AA012427			
BB	AA012427			
BC	AA012427			
BD	AA012427			
BE	AA012427			
BF	AA012427			
BG	AA012427			
BH	AA012427			
BI	AA012427			
BJ	AA012427			
BK	AA012427			
BL	AA012427			
BM	AA012427			
BN	AA012427			
BO	AA012427			
BP	AA012427			
BQ	AA012427			
BR	AA012427			
BS	AA012427			
BT	AA012427			
BU	AA012427			
BV	AA012427			
BW	AA012427			
BX	AA012427			
BY	AA012427			
BZ	AA012427			
CA	AA012427			
CB	AA012427			
CC	AA012427			
CD	AA012427			
CE	AA012427			
CF	AA012427			
CG	AA012427			
CH	AA012427			
CI	AA012427			
CJ	AA012427			
CK	AA012427			
CL	AA012427			
CM	AA012427			
CN	AA012427			
CO	AA012427			
CP	AA012427			
CQ	AA012427			
CR	AA012427			
CS	AA012427			
CT	AA012427			
CU	AA012427			
CV	AA012427			
CW	AA012427			
CX	AA012427			
CY	AA012427			
CZ	AA012427			
DA	AA012427			
DB	AA012427			
DC	AA012427			
DD	AA012427			
DE	AA012427			
DF	AA012427			
DG	AA012427			
DH	AA012427			
DI	AA012427			
DJ	AA012427			
DK	AA012427			
DL	AA012427			
DM	AA012427			
DN	AA012427			
DO	AA012427			
DP	AA012427			
DQ	AA012427			
DR	AA012427			
DS	AA012427			
DT	AA012427			
DU	AA012427			

[illegible]

RESULT	7
AB50965	
ID	AB50965 standard; Protein: 716 AA.
XX	
AC	AB50965;
XX	
DT	21-MAR-2001 (first entry)
XX	
DE	Human PRO1338 protein.
XX	
KW	Human; PRO; cytosolic; neuroprotective; respiratory general;
KW	antiinflammatory; antiangiogenic; immunosuppressive; immunostimulant;
KW	PRO agonist; cancer; inflammatory disorder; immunological disorder.
XX	
OS	Homo sapiens.
XX	
PN	WO200073348-A2.
XX	
PD	07-DEC-2000.
XX	
PF	30-MAY-2000; 2000WO-US14941.
XX	
PR	02-JUN-1999; 99WO-US12252.
PR	22-JUN-1999; 99US-0140650.
PR	23-JUN-1999; 99US-0141037.
PR	20-JUL-1999; 99US-0144758.
PR	01-SEP-1999; 99WO-US20111.
PR	08-SEP-1999; 99WO-US20594.
PR	29-OCT-1999; 99US-0162506.
PR	30-NOV-1999; 99WO-US28313.
PR	01-DEC-1999; 99WO-US28634.
PR	02-DEC-1999; 99WO-US28551.
PR	16-DEC-1999; 99WO-US30095.
PR	20-DEC-1999; 99WO-US30999.
PR	06-JAN-2000; 2000WO-US00376.
PR	11-FEB-2000; 2000WO-US03565.
PR	18-FEB-2000; 2000WO-US04341.
PR	18-FEB-2000; 2000WO-US04342.
PR	02-MAR-2000; 2000WO-US05841.
PR	03-MAR-2000; 2000US-0187202.
PR	10-MAR-2000; 2000WO-US06319.

[illegible]

The present sequence represents a human secreted polypeptide. The secreted polypeptides and polynucleotides are useful for treating nervous and muscular diseases, for inhibiting tumour formation and metastasis, for treating gastrointestinal ulceration, for preventing and treating diseases in spinal cord, thyroid gland, ovary, prostate, renal gland, small intestine, heart, trachea, thymus, lymph node, muscular system and colon, for treating lipase deficiency in cystic fibrosis and pancreatitis, for treating undesirable clot formation such as myocardial infarction, during angioplasty and all surgical procedures that require decreased blood clot formation, for treating liver diseases, coagulation disorders and microbial diseases, for treating immune disorders, for treating inflammation and transplant rejection, for enhancing bone thickness and increasing bone density, for reducing the loss of essential ferroxidases, for suppressing apoptosis, and for regulating vascular smooth cell proliferation. They may also be used as vaccines.

QY	29	CPFOCACQIRPMWYPRSSRYRATVYDNCNDLFTAVPALPAGVQTLLETSNSYTVR	VDOS	87
Db	32	CPQLVCCELRPFQPSRYRATVYDNCNDLFTAVPALPAGVQTLLETSNSYTVR	VDOS	89
QY	88	ELGYLANTLEITDLSQNSFSFARDCCDFHALPOLLSLHBEHQULRLEBHSFAGLASIQELV		147
Db	90	ELGYLANTLEITDLSQNSFSFARDCCDFHALPOLLSLHBEHQULRLEBHSFAGLASIQELV		149
QY	148	LHNDQUTYRPAFAFGSLNLRLEHNSULRAIDSRMEFEMLPNLEIIMIGNKVADILDM		207
Db	150	LHNDQUTYRPAFAFGSLNLRLEHNSULRAIDSRMEFEMLPNLEIIMIGNKVADILDM		209
QY	208	NFRPLANTSLVLAGMNLREISDVALBELQSLSESIFPDNDLAPVPRRALREOVPGKFLD		267
Db	210	NFRPLANTSLVLAGMNLREISDVALBELQSLSESIFPDNDLAPVPRRALREOVPGKFLD		268
QY	268	LKNNLPQRYGPDPEANMLHLKELGJLNNEEELVSDIKFALVNLPELTKLDTITNNRSLFIH		327
Db	270	LKNNLPQRYGPDPEANMLHLKELGJLNNEEELVSDIKFALVNLPELTKLDTITNNRSLFIH		329
QY	328	PRAFHHLPMOEFLEMLNNAALSMLHQOYTESIPNIOEVLHGNPLRCPDVIHMAVAATGRV		387
Db	330	PRAFHHLPMOEFLEMLNNAALSMLHQOYTESIPNIOEVLHGNPLRCPDVIHMAVAATGRV		389
QY	388	RFIEPOSTLCAAPPDLQRLVREVEFRREMTDHCPLISPSRFPSSLOYAEGSEWYLHCRA		447
Db	390	RFIEPOSTLCAAPPDLQRLVREVEFRREMTDHCPLISPSRFPSSLOYAEGSEWYLHCRA		449
QY	448	LAEPEEITWYPRAGLRRLPRPAHAGRRYRYPPEGTLELRVAEEAGITVCAQNMVAGADT		507
Db	450	LAEPEEITWYPRAGLRRLPRPAHAGRRYRYPPEGTLELRVAEEAGITVCAQNMVAGADT		509
QY	508	KTYSVVYVGRALLQPGHDEQSGLELRVQETHRPNLLSWTPRPNTVSTYUMLTSSNS	SLRG	566
Db	510	KTYSVVYVGRALLQPGHDEQSGLELRVQETHRPNLLSWTPRPNTVSTYUMLTSSNS	SLRG	568
QY	567	QGANFALRLPRGTHSYNTRLRLOATEYUWACLOVAFADANTOLACWAAFTKCA		616
Db	569	QGANFALRLPRGTHSYNTRLRLOATEYUWACLOVAFADANTOLACWAAFTKCA		618
QY	619	TSCHRALGDRPGLIATLALAV	639	
Db	626	QETSTALAAVMGSMFAVSLASLAV	650	

RESULT	9
AA#0376	
ID	AA#0376 standard; Protein; 716 AA.
XX	
AC	
XX	AA#0376;
DT	
XX	22-OCT-2001 (first entry)
DE	
XX	Human polypeptide SEQ ID NO 3521.
XX	
KW	Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer
KW	peripheral nervous system; neuropathy; central nervous system; CNS;
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW	chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW	leukaemia.
XX	
OS	
XX	Homo sapiens.
XX	
PN	MO200153312-A1.

PD 26-JUL-2001.
XX 26-DEC-2000; 2000WO-US34263.
PF 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
FA (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
DR N-PSDB: AAI59332.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
PS
PS Example 6: SEQ ID NO 3521: 10078bp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAI38642-AAI42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
CC
CC
XX Sequence 716 AA;
SQ

Query Match 44.8%; Score 1672; DB 22; Length 716;
Best Local Similarity 53.1%; Pred. No. 2.5e-125;
Matches 332; Conservative 105; Mismatches 168; Indels 20; Gaps 5;

QY 29 CPGCACQIRPWYPRSSYREATVDCNDLFLAVPPALPAGTQTLLQSNSTIVR-VDOS 87
DB 32 cpqlcveclrpwftpgstyrealtvdcndlrilrtrpnlssdcqylllgsnmlaktvd-- 89
QY 88 ELGTIANTELDLQNSFSQARCDERHALPQLSLHEENQRLREDDHSGAGASIOELX 147
DB 90 elqqlfnlteldfsqnmfnlkevglanlcltllhleeqqltemtdycldslngely 149
QY 148 LNHQLYRIAPRAFSGLSNLRLHLNSLRLAIDSRFEMIPNLDEIMIGNVVDAILDM 207
DB 150 lnhqlyriaprafsghlslnlrlhlhlnsnrlklvidsrfdstpyldlmisnpanpyldlm 209
QY 208 NFRPLANRSLVLAGMNIRESIDYALBGLQSLSFYDQNLARVRRALEQVGLKFD 267
DB 210 nfrplanrslvlagmnlrlsldlrlhlnsnrlklvidsrfdstpyldlmisnpanpyldlm 269
QY 268 LKNNPLOWVGPDPANMLHLKELGNMMEELVSDKFAVLNLEPRLTLDITNNPRLSEFH 327
DB 270 lknnplovwgdpdpanmlhlkeltgnmmeelvsdkfavl nleprltlditnnprlsef 329
QY 328 PRAFNHLPQMETLMLNNALSALHQGTVESLPNLQEVGLGNPRLCDCVIRMANATCTRY 387
DB 330 prafnhlpqmetlmlnnalsalhqgtveslpnlqevglgnp rldcdcvirmanatctry 387

DB 330 rlafswpaleslmlnnalnaayqktveslplnlreislsnplrfcdvhlhnsnkn 389
QY 388 RFIPEOSTLCAEPDLORLPYREVPFREMDDCLPLISPRSPSPQVSGSMVLHCRA 447
DB 390 rfipemlcmfcamppeyghykvkelyldqsdsegclpmshsfnrlnvdlgtvtfldcra 449
QY 448 LAEPPEIYWTYPAGLRLTPAHAGRRYRYPESTLEIRRYTAEEAGLYTCVAQNLVGADT 507
DB 450 maeppeiywtpipgnklvtelsdkylissegtleisnltqedsgrylcvagngadt 509
QY 508 KTVSVYVGRALLQGRDEGGLELRVQETHPHYHLLISMTWPPPTVSTNLTWSSAS-SLKG 566
DB 510 rvaltkvngcll----qgtqylklyvqcleslsvskvsnvmsnlskwsaakmklion 565
QY 567 QGATALARLPGRGTHSYNITRLQATEYACLOVAFADAHTQACVWARKTA----- 618
DB 566 phlytaerpvdvheynlthlqpsdyevclsvsnlhqgqkscvntcknaafavdlad 625
QY 619 ----TSCHRALGDRPGLAIIALAV 639
DB 626 qetstalaavngsmfavislaslav 650

RESULT 10
AAB42513
ID AAB42513 standard; Protein: 705 AA.
XX
XX AAB42513;
XX
XX 08-FEB-2001 (first entry)
XX
XX Human ORFX ORF2277 polypeptide sequence SPQ ID NO:4554.
DE
XX
XX Human: open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerable; antiproliferative; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antihypertoid;
KW antinaeemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
OS Homo sapiens.
XX
XX WO200058473-A2.
PN
XX
XX 05-OCT-2000.
PD
XX
XX 31-MAR-2000; 2000WO-US08621.
PE
XX
XX 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
PA
XX
XX Shimkets RA, Leach M;
PI WPI: 2000-602362/57.
DR N-PSDB: AAC76722.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 11: Page 3752-3754; 5507pp; English.


```
FT /note= "this region is similar to fibronectin type  
FT III repeat"  
FT Region  
FT 521..604  
FT /note= "this region is similar to fibronectin type  
FT III repeat"  
FT Binding-site  
FT 277..279  
FT /label= RGD motif  
FT /note= "specific tripeptide sequence important for  
FT cell-cell adhesion and cell-matrix adhesion"  
PN WO200009690-A1.  
PD 24-FEB-2000.  
XX  
XX  
XX 09-AUG-1999; 99WO-US17997.  
XX  
XX 10-AUG-1998; 98US-0131648.  
XX  
XX (INCY-) INCYTE PHARM INC.  
PI Hillman JJ, Yue H, Corley NC, Guegler KJ, Patterson C;  
XX WPI; 2000-224335/19.  
DR N-PSDB; AA50941.  
XX  
XX New human extracellular adhesive polypeptide and polynucleotide useful  
PT for diagnosis, prevention and treatment of cancer, immune disorders and  
PT developmental disorders -  
XX  
XX  
XX Claim 1; Pages 58-59; 67pp; English.  
XX  
XX The present sequence is an extracellular adhesive  
CC protein, EXADH2 which was first identified in Incyte clone 2687731  
CC obtained from LUNGNOT23 cDNA library. The cDNA library was constructed  
CC using RNA isolated from left lobe lung tissue removed from a 58 year old  
CC Caucasian male afflicted with osteosarcoma. This sequence is highly  
CC expressed in libraries derived from lymphocytes.  
CC The present sequence is useful in the diagnosis, treatment and prevention  
CC of disorders associated with abnormal expression of EXADH particularly  
CC cancers such as leukaemia and melanoma, immune disorders such as  
CC rheumatoid arthritis, asthma and atherosclerosis, and developmental  
CC disorders such as anaemia, epilepsy, Cushing's syndrome and any disorder  
CC associated with cell growth, differentiation, embryogenesis and  
CC morphogenesis. EXADH, its catalytic or immunogenic fragments are also  
CC useful for drug screening.  
XX  
XX  
SQ Sequence 708 AA;  
  
Query Match 44.4%; Score 1656.5; DB 21; Length 708;  
Best Local Similarity 47.2%; Pred. No. 4; Je-124;  
Matches 333; Conservative 109; Mismatches 217; Indels 47; Gaps 5;  
  
QY 17 TATVVPVPHVHPQPCACQIRPWTPTPRSSREATTVDNCNLFPAVPALPACTOTLL 76  
DB 17 tTlVgvdKvdcprlcceirpWftrsiymeastvdcndlglltParlparqtlll 76  
QY 77 QGNSIVRVQDSEGLVLANITLDELDSQNSFSPARDCDFHALPQLSLHLENQTLREDS 136  
DB 77 qGnIakIeys-tdfpVnltGldIsqnlSsvtnIvKkmpqIIsVyleekIteIpeKc 135  
QY 137 FAGLSLQELVYNHQLRIAPRAFSGLSNLLRLHLSNLLRALDSRMFEKLPVLELMI 196  
DB 136 lSeIenIgeLyInhnlIstIsPgafIgInhllrIhIsnrlqmlnSkwfdalPnlleImI 195  
QY 197 GGNKVDAIILDMNFRPLANLRSIVLAGMNLREISYALEGSLDSLSFYDNOQLARVBRRA 256  
DB 196 gEnpIrlrIdmIfkPlIrlsrlvIagInlIeIpnaIvglenIsIsIfydrIlIkvPhva 255  
QY 257 LRGVGLKFLDLINKNPLORVGPGDFANMLHLKEIGLINMMEVLSIDKFAVLVNLDELKLD 316  
DB 256 lqkvvnlkFlIdInkPnlIrIrgdIsImIhIkEIgInMpelIsIdIsIavdnIpdIrlKle 315
```

```
QY 317 ITNPRLSFIHPRAFHHLPMQETMLNNNALSAHQQTVESLPNLOEVGIGHNPJRCDCV 376  
DB 316 atnprlSylhpnafIfkPlkIesImInsaIsalYhgtIesIpnIkIeshsnpIrdcV 375  
QY 377 IRMANATGTRVRFIEPOSTLCAEPPDQRLPVREKVPREMTDNCPLPISRSPPSLQVA 436  
DB 376 IrmmnmkInlrImepdIscvdpPeTqgqvIvghIrdmmeIcIplIapesIpsInIve 435  
QY 437 SGESKVLHCRALAEPEPIYVWTPAGLRLTPAHAGRRVRYPEGTLELRVTAEEAGLYT 496  
DB 436 agsYsfncraTaepgepIlywIpsgqkIlnrItldkIffYhsagIldIngvtrpKegIyt 495  
QY 497 CYAQLVCAADTKTVSVVYGRALLDPCRBEGCGELFRLQETHPRHILLSWTPNTVSTNL 556  
DB 496 cIatnlvgadIkSvmlKvdsfIpq---dnngsInIkIrIdqansvIvswKasskIlIksv 552  
QY 557 TWSSASLSRGGCATVLAIRLPRGTSYNTIRLQATEFWACLOVAFDAHQQLCVARTR 616  
DB 553 kvtafvKtenshaagsarIpsdvkvynlthnpsIegkIdIdIplYqkrIkKcVnvtK 612  
QY 617 -----EATSchRALGDRPGLIATIALAVLIANGLAHSLTGCPKRGVGR- 663  
DB 613 gIhpqkYeknntIImacIlgIlgIyIclIscl-----spemncdYgns 660  
QY 664 -----PLPPAWAFWGSAPSvRYVSAPLVLPWN 691  
DB 661 yvrynIqkptalgeIypIlnIweagkeksIkvatvIglpIn 706  
  
RESULT 14  
AAM39309  
ID AAM39309 standard; Protein: 708 AA.  
XX  
XX AAM39309;  
AC  
XX  
XX 22-OCT-2001 (first entry)  
DI  
XX  
XX Human polypeptide SEQ ID NO 2454.  
DE  
XX  
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; hemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW Leukaemia.  
XX  
XX  
XX Homo sapiens.  
OS  
XX  
XX  
XX WO200153312-A1.  
PN  
XX  
XX 26-JUL-2001.  
PD  
XX  
XX 26-DEC-2000; 2000WO-US34263.  
PF  
XX  
XX 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao Qa, Zhou P, Goodrich R, Drmanac RT;  
XX  
XX WPI; 2001-442253/47.  
DR N-PSDB; AA158465.  
XX  
XX Novel nucleic acids and polypeptides, useful for treating disorders  
PT
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 31, 2002, 14:34:58 ; Search time 24.35 Seconds
(without alignments)
2813.622 Million cell updates/sec

Title: US-09-905-088a-245

Perfect score: 3732
Sequence: 1 MRLVAPLLAVLWAGATATV.....RKLPRSSRGETLLPPLLSONS 7131

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1649.5	44.2	707	2 JC7763	neuronal leucine-r
2	442	11.8	1091	2 A58532	glial cell membran
3	407.5	10.9	605	2 A41915	insulin-like growt
4	397.5	10.7	605	2 JC5239	insulin-like growt
5	381.5	10.2	789	2 T28714	hypothetical prote
6	381.5	10.2	1355	2 T28715	hypothetical prote
7	368	9.9	907	2 U60193	G protein-coupled
8	360.5	9.7	1531	2 T42218	slit-1 protein hom
9	360	9.6	907	2 JE0176	orphan G protein-c
10	346	9.3	1523	2 T13953	MEG5 protein - ra
11	344	9.2	603	2 JC1282	insulin-like growt
12	342.5	9.2	603	2 JC6128	insulin-like growt
13	324	8.7	1469	2 B36655	slit protein 2 pre
14	324	8.7	1480	2 A36655	slit protein 1 pre
15	321	8.6	560	2 A60164	platelet membrane
16	317.5	8.5	702	2 T21148	hypothetical prote
17	302	8.1	1025	2 T42626	secreted leucine-r
18	299.5	8.0	361	2 A53860	chondroadherin pre
19	298	8.0	536	2 A34901	lysine carboxypept
20	288	8.0	680	2 T19939	hypothetical prote
21	285	7.9	738	2 T19938	hypothetical prote
22	290.5	7.8	1119	2 AD1822	leucine-rich repea
23	289.5	7.8	594	2 T23841	hypothetical prote
24	287.5	7.7	458	2 T19941	hypothetical prote
25	286.5	7.7	610	2 T23395	hypothetical prote
26	285.5	7.7	610	2 T23836	hypothetical prote
27	285.5	7.7	1385	2 T13887	tlr protein - frui
28	283	7.6	1066	2 T15864	hypothetical prote
29	283	7.6	1389	2 T13852	gene wheeler prote

30	280.5	7.5	662	2 S42799	garp precursor - h
31	276.5	7.4	357	2 S24317	decorin precursor
32	274.5	7.4	575	2 T29972	hypothetical prote
33	273	7.3	1112	2 T10504	disease resistance
34	272	7.3	1134	1 A29944	chaptlin precursor
35	270.5	7.2	420	2 A53531	oncietal trophobl
36	268	7.2	354	2 A53454	decorin precursor
37	268	7.2	360	2 S06280	decorin precursor
38	266	7.1	1097	2 A29943	coll protein precu
39	266	7.1	1535	2 S46224	peroxidasin - frui
40	263	7.0	1016	2 T30553	disease resistance
41	262.5	7.0	562	2 T34319	hypothetical prote
42	262.5	7.0	1051	2 T13174	gp150 protein - fr
43	261	7.0	359	1 NBH0C8	decorin precursor
44	261	7.0	360	2 I47020	decorin - rabbit
45	258	6.9	354	2 S29145	decorin precursor

ALIGNMENTS

RESULT 1
JC7763
neuronal leucine-rich repeat protein-3 - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 01-Feb-2002 #sequence-revision 01-Feb-2002 #text-change 01-Feb-2002
C:Accession: JC7763
R:Fukunishi, K.; Matsuo, Y.; Kitano, C.; Kuchino, Y.; Tsuda, H.
Biochem. Biophys. Res. Commun. 287, 257-263, 2001
A:Title: Rat neuronal leucine-rich repeat protein-3: Cloning and regulation of the ge
A:Reference number: JC7763; PMID:11549284
A:Contents: Fibrosarcoma cells
A:Accession: JC7763
A:Molecule type: mRNA
A:Residues: 1-707 <FW>
A:Cross-references: GB:AF291437
C:Comment: This protein, a new member of the neuronal leucine-rich repeat protein fam
in protein-protein interaction and functions as a cell adhesion molecule or soluble I
C:Genetics:
A:Gene: nlrr-3
C:Keywords: cell adhesion

Query Match	44.2%	Score 1649.5;	DB 2;	Length 707;
Best Local Similarity	47.5%	Pred. No. 4.5e-105;		
Matches 337;	Conservative 112;	Mismatches 229;	Indels 31;	Gaps 7;
QY	6	APLLAWAG--ATATPVVPMHVPCCPACQIRPWYTPRSYREATVDCNDLFTAV	63	
Db	4	APLDIHVLGLDALTALVYQAGDKKVDPCQICEIRPWTFRSIVWASTVDCNDLGLNF	63	
QY	64	PPALPACTGTLILQNSIVRVDSGLGYLANTELDSONSFSDARDCDFALPQLSLH	123	
Db	64	PARLPACTGTLILQNTNINAKIENS-TDFPVNLTGDLSSNNLSSTNTINVKMSQLLSVY	122	
QY	124	LEENQLRLDHSFAGLASLOEYLNNHQLYRIAPRSGSLNLRLLNSLRAISR	183	
Db	123	LEENKLELEPKLIGSLNQLIYVHNLISAGATVGLNLRLLNSLRAISR	182	
QY	184	WFEMLPNLEILMTGKNKVDAILDMNFRPLANRSLVLAGMNLRETSYDALGLQSLSL	243	
Db	183	WFEMLPNLEILMTGKNKVDAILDMNFRPLANRSLVLAGMNLRETSYDALGLQSLSL	242	
QY	244	FYDNLARVPRRLAEQVPGIKPLDKNPLORVSGDFANMILKEIGLNMEELVSIDK	303	
Db	243	FYDNLARVPRRLAEQVPGIKPLDKNPLNIRRGDFSNNILKEIGLNMEELVSIDS	302	
QY	304	FALVNEBELTKLDTNNPRLSPFTHRAFHRLPOMETLMLNNVALSALHQQTVESLPIQ	363	
Db	303	LAVNLPDLKRIKATANNPRLSYTHRAFHRLPKLSIMLNSALSYHGTIESLPIKE	362	
QY	364	VGLHGNIRDCVYIRANATGTRVRIEFGSLCAEPDGLQRLPVRKVPFEMMDHCLPL	423	


```

Db 178 -----PVOA-----FRSLAQLAMTLANKTHHTADYAFGLMSLTV 213
Oy 241 SLSPYDNOIARVPARRALEOVPLKFLDLINKNPLQVRGPDPAFMLHKEIGLNNEEYVS 300
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 214 VHLHNNRHSIGKCKCFDGLHLETLIDLNNYNDLDF-FLAIKTLNLELGFHS-NNIRS 271
Oy 301 IOKFALVNIPELTKLDITNNPRLSFIHPRAEHHLPMQETMLNNA----- 346
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 272 IPERAFVGNPSLITTHFYDNP-IGFVGSAPQHLPELRLTLNGASHTERPHLTGATL 330
Oy 347 -----LSALHOQTVESLPNLOEVGLHGN 369
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 331 ESLTLTGAKITSLPQAVCDPLNQLVLDLSYN 362

RESULT 8
T42218
sift-1 protein homolog - rat
N:Alternate names: MEGF4 protein
C:Species: Rattus norvegicus (Norway rat)
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000
C:Accession: T42218
R:Nakajima, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
A:Reference number: Z14126; MUID:98360089
A:Accession: T42218
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1531 <NAK>
A:Cross-references: EMBL:AB011350; NID:g3449289; PIDD:BAA32460.1; PID:g3449290
A:Experimental source: strain Sprague-Dawley; brain
C:Genetics:
A:Gene: MEGF4

```

```

Query Match 9.7%; Score 360.5; DB 2; Length 131;
Best local similarity 23.6%; Pred. No. 3e-16;
Matches 136; Conservative 76; Mismatches 216; Indels 149; Gaps 17;

```

```

Oy 8 LLL---AWVAGTATVPVPMHVPCCQACQIRWTPRSSYREATTVDNDLFTAVP 64
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 20 LLLMAARLGLTA-----CPALCTC-----TGTVDDCHGGLQALIP 56
Oy 65 PALPAGTOTLLQNSIVRVDOSELGYLANTELDLSONSFSDARDCDFHALPOLLISL 124
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 57 KNIPNTERLELNGNNITRIHKNDFAGIKQRLVQLQMENQIGAVRGAFDDMKELERL 116
Oy 125 EBNQUTRLDHSFAGLASLQELYNHNOYLRIAPRAGSLNLRHLNLSMLRAIDSRW 184
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 117 NKNQLOVPELLEFQNNQALSRDLSENQLQVPRKAFRGATVDLKNLQIDKNQISCEGA 176
Oy 185 FEMLEPNEILMIGKKNVAIIDMFRPLANLRSL----- 218
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 177 FRALGLLEVLITNNNNITTTIPVSSFNMPKILKTRFLHSNHLFCCHLAMLSQLRQPTI 236
Oy 219 -----VLAGNNLEI--SDYALEGLQSLBSLSFYDNOIARVPR----- 254
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 237 GLFTQCGSPASLRGLNVAEYQKSEFSCG-----GQEAQVPRCTLSGSGCPAMC 286
Oy 255 -----RALEQVVG-----LKFLLDLKNKPLQRYGPDGFANMLHLKELGLNNEEL 298
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 287 SCNSGIVDCRGKGLTAIPANLPEMTETRLRLNGIKSIPGAFSPYRKLRIRIDISN-NOI 345
Oy 299 VSIDKFAVLNPELTKL-----DITNNPR-----LSFTHPRAPHILP 335
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 346 AEIADAPQGLSLNSLVLYGNKKTIDLPFGVFGGITYTLQILLANANKINCIRPRAFQDLQ 405
Oy 336 QMETLNNNNLSALHQTVESLPNLOEVGLHGNPRLDCVYIRMANATGTRVRETFPOST 395
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 406 NLSLSTLVNKKIOSLAKGTFTSLRAIQTLHLAQNPFIDCMLKVL-ADFLTNPIETTTGA 464
Oy 396 LCAEPDIDQRLPVRVPRREM-----TDHCLPLISPSFPSSLOVAGSESVLHCRALAE 450

```

```

Db 465 RCASPRRLANKRIGQIKSKFKCSAKQYFI--GTEDYHLNSECSDVACPRHKC--- 519
Oy 451 PEPETVYTAGLRLT--PAHAGRRVYPEGLELR 485
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 520 ---EASVYECSGIKLSKIPER-----IPQSTELR 546

RESULT 9
JE0176
orphan G protein-coupled receptor precursor - human
C:Species: Homo sapiens (man)
C:Date: 03-Jul-1998 #sequence_revision 10-Jul-1998 #text_change 21-Jul-2000
C:Accession: JE0176
R:McDonald, T.; Wang, R.; Bailey, W.; Xie, G.; Chen, F.; Caskey, C.T.; Liu, Q.
Biochem. Biophys. Res. Commun. 247, 266-270, 1998
A:Title: Identification and cloning of an orphan G protein-coupled receptor of the gl
A:Reference number: JE0176; MUID:98308104
A:Accession: JE0176
A:Molecule type: mRNA
A:Residues: 1-907 <MCD>
A:Cross-references: GB:AF062006; NID:g3366801; PIDD:AAC28019.1; PID:g3366802
A:Comment: This protein is a receptor for a novel class of glycoprotein ligands.
C:Genetics:
A:Gene: HG38
A:Map position: 12q22-23
F:1-21/Domain: signal sequence #status predicted <SIG>
F:562-583/Domain: transmembrane #status predicted <TM1>
F:594-616/Domain: transmembrane #status predicted <TM2>
F:639-660/Domain: transmembrane #status predicted <TM3>
F:681-701/Domain: transmembrane #status predicted <TM4>
F:725-744/Domain: transmembrane #status predicted <TM5>
F:768-791/Domain: transmembrane #status predicted <TM6>
F:803-824/Domain: transmembrane #status predicted <TM7>

```

```

Query Match 9.6%; Score 360; DB 2; Length 907;
Best local similarity 27.4%; Pred. No. 1.5e-16;
Matches 134; Conservative 70; Mismatches 189; Indels 96; Gaps 14;

```

```

Oy 3 LVAPELLAWAGTATVPVPMHVPCCQACQIRWTPRSSYREATTVDNDLFTAVP 62
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 9 LLSLPVLLQATGSSPSRSGVLLR-GCPTHGCE-----PDGRMLLRVDCSDGLSE 59
Oy 63 VPPALPAGTOTLLQNSIVRVDOSELGYLANTELDLSONSFSDARDCDFHALPOLLISL 122
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 60 LPSLVSFTSYLDLSMNNISQGLPRLPSRLTEBELRLAQNATLYTPKCAFTGLYSLKYL 119
Oy 123 HLEBNQUTRLDHSFAGLASLQELYNHNOYLRIAPRAGSLNLRHLNLSMLRAIDS 182
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 120 MLONNQLRHVPBALQMLRSLSQSLRLDANHSYVPPSCFGLSHRLMLDNLATLEI-- 177
Oy 183 RWEPMLEPNEILMIGKKNVAIIDMFRPLANLRSVLVLAGMNLREISDVALEBQSLSE 242
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 178 -----PVOA-----FRSLAQLAMTLANKTHHTADYAFGLMSLTV 215
Oy 243 SFYDNOIARVPARRALEOVPLKFLDLINKNPLQVRGPDPAFMLHKEIGLNNEEYVS 302
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 216 HLNHNNRHSIGKCKCFDGLHLETLIDLNNYNDLDF-FLAIKTLNLELGFHS-NNIRSIP 273
Oy 303 KFAVLNIPPELTKLDITNNPRLSFIHPRAEHHLPMQETMLNNA----- 346
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 274 EKAFVGNPSLITTHFYDNP-IGFVGSAPQHLPELRLTLNGASQITTEPPDLGTANLES 332
Oy 347 -----LSALHOQTVESLPNLOEVGLHGNPRLDCVYIRMANATGTRVRETF-POSTIC 397
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 333 LTLTGAQISSLPQVGNQPLNQLVLDLSYN-----LLEDDPSFSVC 373
Oy 398 AEPDIDQRLPVRVPRREM-TDHCPLPISPSFPSSLOVAGSESVLHCRALAEPEPELY 456
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 374 QK---LQKIDLRNLEIYELKVDTFQQLLSLR-----SLNLANMKIALITH-----PNAF 418
Oy 457 WTPPAGRL 465

```


R:Boisclair, Y.R.; Seto, D.; Hsieh, S.; Hurst, K.R.; Ooi, G.T.
Proc. Natl. Acad. Sci. U.S.A. 93, 10028-10033, 1996
A:Title: Organization and chromosomal localization of the gene encoding the mouse acid 1
A:Reference number: J06128; MIMD:96413591
A:Accession: J06128
A:Molecule type: DNA
A:Residues: 1-603 <BOI>
A:Cross-references: GB:066900; NID:g1621612; PIDN:AB17270.1; PID:g1621613
C:Comment: This protein is a serum protein and it is of the ternary complex in the phys
A:Gene: als
A:Map position: 17

Query Match 9.2%; Score 342.5; DB 2; length 603;
Best Local Similarity 22.0%; Pred. No. 1.3e-15;
Matches 139; Conservative 69; Mismatches 190; Indels 233; Gaps 15;

```
QY 8 LLLMVA-----GATATPVVPMVHPDPOCAQIRWYPRSSYREATVDGN 56
  || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 13 LLAFWALGPCYLOGTDPGASADAE---GRQCVTCTCY-----DDYDELVSFCS 61
QY 57 DLFLVAPPAAPAGTQTLLQSNIVY-----DOSELY 91
  || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 62 SRNLQLPDGIPTSTRAIMLDGNLSSIPSAFQNLSDFLMLQGSWLSLEPQALLG- 120
QY 92 LANLELDLSONSFSDARDCDFHLPOLLSLHEENQTLFE----- 133
  || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 LQNLHLHLERLLRLSLAAGLFRHTPSLASLNNLGLREGLRGLSHLMDLNGM 180
QY 134 -----DHSFAGLSLOELYNHOLYRIAPRAFSGLNLRHLNSLRALDSWFEM 187
  || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 SLVLPDVFQGLGHELYLAGNKLTYPDALLCGELRELDLSRNALRYKANVEIH 240
QY 188 LPNELLIMGNKYDAI-----LDNM----- 208
  || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 LPRLQKLTLDRLITVAAPRAFLGMKALRWLDLSHNRAVLLEDTPRGLLGHVRLAHN 300
QY 209 -----FRPLANKSLVLAGNLRLEISDYLEGLOSSESIFYNOIARY----- 252
  || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 AITSRRPTFRFDHLELDLQGNHRIQRLGKREGIGOLEVLTINQIHVKVAFEG 360
QY 233 -----PRALBOYFGIKFLDLKNPDLQRYGQDDPANNMLHKLGLNNM 295
  || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 361 LFNVAVMNLSONCLRSLEHVEOGIGRLHSLHESCLGRIRHTFAGLSGLRRLFLRD- 419
QY 296 EELVSDKFAVLNPELTKLDITNNPRLSFTHPRAFHLLPOMETLMNNALSALHQ--- 352
  || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 420 NSISIEEOSIAGLSELELDLITAN-QLTHLPRLPQGLQLELLSNQOLMISEDYL 478
QY 353 -----QVESLPNLOEVLGNPIR 372
  || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 479 GLPQAFWLDLSHNLETPAGELFSGLRYLNLNNNSLQTFVPOGELRLMLDNPMD 538
QY 373 CDCVIR-----WANATGRVRETEP-----QSLCAEPDQLRLVREVPFR 414
  || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 539 CSCPALKALRDALQNPVYPRFVOTVCEGDCCOPVYTYNNITCAGPANSGLDLRL- 595
QY 415 EMTDHCPLISPRSPSLQVASESMVLC 445
  || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 596 -----SETLFLVHC 603
```

RESULT 13
B3665
Silt protein 2 precursor - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C:Date: 30-Apr-1991 #sequence-revision 30-Apr-1991 #text-change 19-May-2000
C:Accession: B3665
R:Rothenberg, J.M.; Jacobs, J.R.; Goodman, C.S.; Artavanis-Tsakonas, S.
Genes Dev. 4, 2169-2187, 1990
A:Title: silt: an extracellular protein necessary for development of midline glia and cc
A:Reference number: A3665; MIMD:91093665

A:Accession: B3665
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1469 <ROF>
A:Cross-references: GB:X53959
C:Genetics:
A:Gene: FlyBase:sil
A:Cross-references: FlyBase:FBgn0003425
C:Superfamily: unassigned EGF-related proteins; EGF homology; leucine-rich alpha-2-gl
F:66-91/Domain: proteoglycan amino-terminal homology <PAH1>
F:101-124/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR1>
F:125-148/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2>
F:149-172/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>
F:173-196/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>
F:197-220/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5>
F:228-243/Domain: proteoglycan carboxyl-terminal homology <PCS1>
F:288-316/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR6>
F:323-346/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR7>
F:347-370/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR8>
F:371-394/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR9>
F:395-418/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR10>
F:419-442/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR11>
F:450-454/Domain: proteoglycan carboxyl-terminal homology <PCS2>
F:512-537/Domain: proteoglycan amino-terminal homology <PAH2>
F:547-571/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR12>
F:572-595/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR13>
F:596-619/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR14>
F:620-643/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR15>
F:651-685/Domain: proteoglycan carboxyl-terminal homology <PCS3>
F:708-733/Domain: proteoglycan amino-terminal homology <PAH4>
F:743-766/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR16>
F:767-790/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR17>
F:846-890/Domain: proteoglycan carboxyl-terminal homology <PCS4>
F:1028-1061/Domain: EGF homology <EGF1>
F:1068-1099/Domain: EGF homology <EGF2>
F:1115-1148/Domain: EGF homology <EGF3>

Query Match 8.7%; Score 324; DB 2; length 1469;
Best Local Similarity 24.7%; Pred. No. 9e-14;
Matches 103; Conservative 67; Mismatches 169; Indels 78; Gaps 11;

```
QY 29 CPQCAQIRWYPRSSYREATVDCNDLFLVAPPAAPAGTQTLLQSNISVRRDSE 88
  || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 519 CPAMCHC-----EGTVDCIGRRLEIPRDLPLHTTELLLDNDELGRISDG 565
QY 89 LGYLANLELDLSONSFSDARDCDFHLPOLLSLHEENQTLFEHPSFAGLSAQLLY 148
  || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 566 L-----FGRLPHLYNLTKLRNQLTGIEFNAPFGASHIOELQ 602
QY 149 NNNQYRIAPRAFSGLNLRHLNSLRALDSRWFEMLPNELLIMGNKYDAIILDMN 208
  || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 603 GENKIKEISNMFGLQLQKTLNLYDNQISCVMPGSEHNSLTSINLASNPENC--NCH 660
QY 209 FRPLAN-LRSVLVLAGMNLR-----EISDYALEGLOSSE----- 240
  || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 661 LAMRAECYRKSLNGAARGAASKVADQIKQLPHEFCSSENSEGCGDQCPSPCT 720
QY 241 ---SLSFYNQALARVPRRALBOVG--LKPDLANKPDLQRYGQDDPANNMLHKLGLNNM 295
  || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 721 CTGVVACSRNQLEIPR---GIPAEISLEYSENEIEOIHERRIRHLSLRDLDSN- 775
QY 296 EELVSDKFAVLNPELTKLDITNNPRLSFTHPRAFHLLPOMETLMNNALSALHQIV 355
  || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 776 NQITLSNWTYRANLTKLSTLITSTN-KLQCLQRLALSGLNKLNVLSHGNRISMPEGSF 834
QY 356 ESLPNLOEVLGNPIRDCVYRMANATGRVRETEPQSLCAEPDQLRLVREVP 412
  || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 835 EDLKLTHIALGSLNPLVCDGLKWF--SDWIKLIDYVEGIRCAEPDMKDLITLSTP 890
```

RESULT 14
A3665

C:Comment: The amino end of the intact protein is blocked.
C:Comment: This protein is absent in Bernard-Soulier syndrome.

C:Genetics:

A:Gene: GDB:GP5
A:Cross-references: GDB:230236; OMIM:173511

A:Map position: 5pter-5qter

C:Superfamily: leucine-rich alpha-2-glycoprotein repeat homology
C:Keywords: blocked amino end; glycoprotein; platelet; tandem repeat; transmembrane prot

Query Match 8.6%; Score 321; DB 2; Length 560;

Best Local Similarity 26.6%; Pred. No. 3,5e-14;
Matches 141; Conservative 57; Mismatches 227; Indels 106; Gaps 16;

```
OY 9 LLAVVAGATATVPVPMHVPQPCACQIRPWYTPRSSYREATTVDCNDLF-LTAVPAL 67
DB 6 LICAVALGILRAQP-----FPCPPACKC-----VERDAQCSSGDVARISAL-GI 48
OY 68 PAG-TOTLLQSNSTIVVDSELGVLNLTPELDLSONSFSDARCDFFHALPOLSLHLE 126
DB 49 PTNLTHILTFMGSGVQLQSQSFSG-KTVLQRLMISDSHISAVAPGTFSDLIKTLRLSR 107
OY 127 NQTLRLDEHSPAGLASIQELYLNHNQLYRIAPAFSGLSNLRHLHNSNLLRAIDSRWE 186
DB 108 NKITHLPQALDKKVVLEQLFLDHNLALRGIDQNMFKLVNQLQELALNQNLDFLPASLFT 167
OY 187 MLPNLEILMIGGNKVDALIDMNFRLANLRSIVLAGMNLREISDYALEGLQSLSESFYD 246
DB 168 NLENIKLIDLSGNNLTHTLPKGLGAQAKLERLILHSNRLVSLDSGLINSIGALTLELQFHR 227
OY 247 NQLARVPRRALEQVPGIKFLDLNK-----NPLQRVGPGDPA 282
DB 228 NHIRSIAPGAFDRPLPNLSLTLSSRNHLAFLPSALFLHSHNLTLLTFENPLAELPGVLF 287
OY 283 NMLHKEIGLNNMELVSIKFAVLNPELTKDITNNPRLSFIHPRAFHHPQMETML 342
DB 288 EMGGLQELMLNR-TQLRTLPAAPAFRNLSRLRYLGVTLSPRLSALPGAFQIGELQVLA 346
OY 343 NNNALNAL-----HOQ-----T 354
DB 347 HSNGLTALPDGLLGLGKLRQVSIRNRRLALPRALFRNLSSLESVQLDHNQLETLPGDV 406
OY 355 VESLPNLOEVGLHGNPIRDCD---VIRMANATGTRVRETEPOSTLCAEPPDLQRLVRE 410
DB 407 FGALPRLTEVLIGHNSWRCDCGLCPFLGWLRQHLGLVGCGEPPR--CAGPGAHAGLPLWA 464
OY 411 VPFRMTDHCPLISPRSPSLQVAGSESVLCHRALAEPEPEY-WVTP 460
DB 465 LPGAQA--EC---PGPRGPPPPRAADSSSEAPVH-PALAINSSSEPPWVAQP 509
```

Search completed: August 31, 2002, 14:36:40
Job time: 102 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 31, 2002, 14:35:28 ; Search time 13.6 Seconds
(without alignments)
2029.929 Million cell updates/sec

Title: US-09-905-088a-245
Sequence: 3732
1 MRLVAPLLAWAGATATV.....RKLPSSSEGETLPPPLSQNS 713

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40:*
Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3718	99.6	713	1	GACL_HUMAN
2	407.5	10.9	605	1	AUS_HUMAN
3	397.5	10.7	605	1	AUS_PAPHA
4	344	9.2	603	1	AUS_RAT
5	342.5	9.2	603	1	ALS_MOUSE
6	340.5	9.1	567	1	GPV_MOUSE
7	324	8.7	1480	1	SLIT_DROME
8	321	8.6	560	1	GPV_HUMAN
9	314	8.4	905	1	TRP3_MOUSE
10	313.5	8.4	567	1	GPV_RAT
11	298.5	8.0	361	1	CHAD_BOVIN
12	287	8.0	536	1	CBP8_HUMAN
13	287	7.7	904	1	TRP3_HUMAN
14	280.5	7.5	662	1	GARP_HUMAN
15	279	7.5	782	1	CHAO_TRICA
16	276.5	7.4	357	1	PGS2_CHICK
17	272	7.3	1315	1	CHAO_DROME
18	270	7.2	360	1	PGS2_PIG
19	270	7.2	1050	1	TLR7_MOUSE
20	269.5	7.2	1032	1	TLR9_MOUSE
21	268.5	7.2	1032	1	TLR9_HUMAN
22	268	7.2	354	1	PGS2_MOUSE
23	268	7.2	360	1	PGS2_BOVIN
24	266	7.1	1097	1	TOLL_DROME
25	261	7.0	359	1	PGS2_HUMAN
26	261	7.0	977	1	PGS2_RABIT
27	260	7.0	977	1	Y848_HUMAN
28	259	6.9	331	1	PLIB_AGRBL
29	258	6.9	354	1	PGS2_RAT
30	253	6.8	1049	1	TLR7_HUMAN
31	253	6.8	360	1	PGS2_CANFA
32	253	6.8	372	1	PGS1_HORSE
33	252	6.8	1032	1	TLR8_MOUSE

ALIGNMENTS

RESULT	ID	STANDARD	PRT	713 AA
1	GACL_HUMAN			
AC	075325			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	Glioma amplified on chromosome 1 protein precursor.			
GN	GACL			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Glial tumor;			
RX	MEDLINE=96324709; PubMed=9662332;			
RA	Malfoy B., Almeida A., Zhu X.X., Vogt N., Tyagi R., Mulleris M.,			
RA	Dutrillaux A.-M., Dutrillaux B., Ross D., Hanash S.;			
RT	"GACL", a new member of the leucine-rich repeat superfamily on			
RT	chromosome band 1q32.1, is amplified and overexpressed in malignant			
RT	gliomas."			
RL	Oncogene 16:2997-3002(1998).			
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).			
CC	-1- TISSUE SPECIFICITY: OVERAMPLIFIED IN MALIGNANT GLIOMAS.			
CC	-1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.			
CC	-1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.			
CC	-1- SIMILARITY: CONTAINS 11 LEUCINE-RICH REPEATS (LRR).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AF030435; AAC39792.1; -			
DR	MIM; 605492; -			
DR	InterPro; IPR003006; Ig_MHC.			
DR	InterPro; IPR003598; Ig_C2.			
DR	InterPro; IPR001611; LRR.			
DR	InterPro; IPR000483; LRR_Nterm.			
DR	InterPro; IPR000372; LRR_Cterm.			
DR	InterPro; IPR003592; LRR_out.			
DR	InterPro; IPR003591; LRR_typ.			
DR	Pfam; PF00047; Ig_1.			
DR	Pfam; PF00560; LRR_10.			
DR	Pfam; PF01463; LRRCT; 1.			
DR	Pfam; PF01462; LRRNT; 1.			
DR	PRINTS; PR00019; LEURICHRPT.			
DR	SMART; SM00408; ICG2; 1.			
DR	SMART; SM00370; LRR; 6.			
DR	SMART; SM00082; LRRCT; 1.			
DR	SMART; SM00013; LRRNT; 1.			
DR	SMART; SM00369; LRR_TYP; 2.			

34	251.5	6.7	369	1	PGS1_SHEEP	O46390 ovis aries
35	251	6.7	360	1	PGS2_HORSE	O46542 equus caball
36	250.5	6.7	369	1	PGS1_BOVIN	P21809 bos taurus
37	250.5	6.7	682	1	CONN_DROME	O01819 drosophila
38	249	6.7	369	1	PGS1_MOUSE	P28653 mus musculus
39	248.5	6.7	369	1	PGS1_CANFA	O02678 canis fam1
40	248	6.6	369	1	PGS1_RAT	P47853 rattus norv
41	246.5	6.6	368	1	PGS1_HUMAN	P21810 homo sapien
42	246.5	6.6	2026	1	CYAA_YEAST	P08678 saccharomyc
43	242.5	6.5	582	1	SHO2_HUMAN	O9uq13 homo sapien
44	242.5	6.5	582	1	SHO2_MOUSE	O88520 mus musculu
45	241	6.5	342	1	LUM_BOVIN	O05443 bos taurus

KM Immunoglobulin domain; Transmembrane; Glycoprotein; Repeat;
 KW Leucine-rich repeat; Signal.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 713 GLIOMA AMPLIFIED ON CHROMOSOME 1 PROTEIN.
 FT DOMAIN 19 630 EXTRACELLULAR (POTENTIAL).
 FT TRNSMEM 631 651 POTENTIAL.
 FT DOMAIN 652 713 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 92 115 LRR 1.
 FT REPEAT 116 139 LRR 2.
 FT REPEAT 140 163 LRR 3.
 FT REPEAT 165 187 LRR 4.
 FT REPEAT 188 211 LRR 5.
 FT REPEAT 213 235 LRR 6.
 FT REPEAT 236 259 LRR 7.
 FT REPEAT 261 283 LRR 8.
 FT REPEAT 309 333 LRR 9.
 FT REPEAT 334 357 LRR 10.
 FT REPEAT 359 385 LRR 11.
 FT REPEAT 438 504 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 445 497 BY SIMILARITY.
 FT DISULFD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHD 381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHD 555 555 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHD 583 583 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 713 AA; 78798 MW; EC8BC0FD240C9396 CRC64;

Query Match 99.6%; Score 3718; DB 1; Length 713;
 Best Local Similarity 99.7%; Pred. No. 2,1e-261;
 Matches 711; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 MRLVAVILLAWAGATATVPPVPMHVPCCPQACQIRPMYPRSSYREATTVCNDLFL 60
 DB 1 MRLVAVILLAWAGATATVPPVPMHVPCCPQACQIRPMYPRSSYREATTVCNDLFL 60
 OY 61 TAVPALPAGTQTLTQSSNIVRVDOSELGYLANITELDLSONSFSDARCDFFALPOLL 120
 DB 61 TAVPALPAGTQTLTQSSNIVRVDOSELGYLANITELDLSONSFSDARCDFFALPOLL 120
 OY 121 STHLENOITRELDHSEFAGLALSLQELTYLNHNDLYRAPAFSGSLNULRLHNSMLTARI 180
 DB 121 STHLENOITRELDHSEFAGLALSLQELTYLNHNDLYRAPAFSGSLNULRLHNSMLTARI 180
 OY 181 DSRFEMENLNLITLMIGKKNVDAILDMNFRPLANLRSVLAGMNLREISDYALEGLQSLLE 240
 DB 181 DSRFEMENLNLITLMIGKKNVDAILDMNFRPLANLRSVLAGMNLREISDYALEGLQSLLE 240
 OY 241 SLSFYDNOGLARVPRLAEQVPGIKFLDLKNPLQRYGPGDFPANNMLHLKELGINNMEELVS 300
 DB 241 SLSFYDNOGLARVPRLAEQVPGIKFLDLKNPLQRYGPGDFPANNMLHLKELGINNMEELVS 300
 OY 301 IDKFAVLNPELTKLDITNNPRLSFTHRAPFHLPMOETMLNNNALSLAHQOYESLPN 360
 DB 301 IDKFAVLNPELTKLDITNNPRLSFTHRAPFHLPMOETMLNNNALSLAHQOYESLPN 360
 OY 361 LQEVGLHGNPIRCDVYIRMANATGTRVRELEPOSTLCAEPDQLRVREVPFREKTDHC 420
 DB 361 LQEVGLHGNPIRCDVYIRMANATGTRVRELEPOSTLCAEPDQLRVREVPFREKTDHC 420
 OY 421 LPLISRSRPPSLQVAGSGESMVLHCRALAEPEIYTWPRAGRLTPALAGRCRYRPG 480
 DB 421 LPLISRSRPPSLQVAGSGESMVLHCRALAEPEIYTWPRAGRLTPALAGRCRYRPG 480
 OY 481 TLELRRTAEAGLYTCVAKONTLGADTKTVYVVGRRALQOPGRDESGLELRYOETHRNH 540
 DB 481 TLELRRTAEAGLYTCVAKONTLGADTKTVYVVGRRALQOPGRDESGLELRYOETHRNH 540
 OY 541 TLISWTTPPTVSTNLTWSSASSLRQOGATALARLPRGTHSYNITRLLOATEYMACLOVA 600
 DB 541 TLISWTTPPTVSTNLTWSSASSLRQOGATALARLPRGTHSYNITRLLOATEYMACLOVA 600
 OY 601 FAADHTQLACVMAKTEATSCRRALDRPGLIALLAVILLAAAGLAHLGSGPRKGVG 660
 DB 601 FAADHTQLACVMAKTEATSCRRALDRPGLIALLAVILLAAAGLAHLGSGPRKGVG 660

DB 601 FAADHTQLACVMAKTEATSCRRALDRPGLIALLAVILLAAAGLAHLGSGPRKGVG 660
 OY 661 GRRLPPANAFWGSAPSVRYVSAPLVLPNPGKRLPRSEGEFTLLPPLSQNS 713
 DB 661 GRRLPPANAFWGSAPSVRYVSAPLVLPNPGKRLPRSEGEFTLLPPLSQNS 713

RESULT 2
 ALS_HUMAN STANDARD; PRT; 605 AA.
 ID ALS_HUMAN
 AC P35858;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Insulin-like growth factor binding protein complex acid labile chain precursor (ALS).
 GN IGFBP3 OR ALS.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Liver;
 RX MEDLINE=92357025; PubMed=1379671;
 RA Leong S.R., Baxter R.C., Camerato T., Dai J., Wood W.I.;
 RT "Structure and functional expression of the acid-labile subunit of the insulin-like growth factor-binding protein complex.";
 RL Mol. Endocrinol. 6:870-876(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Frankland J.;
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 28-35.
 RX MEDLINE=89308584; PubMed=2473065;
 RA Baxter R.C., Martin J.L., Beniac V.A.;
 RT "High molecular weight insulin-like growth factor binding protein complex. Purification and properties of the acid-labile subunit from human serum.";
 RL J. Biol. Chem. 264:11843-11848(1989).
 CC -1- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS THAT RESULT IN PROTEIN COMPLEXES. RECEPTOR-LIGAND BINDING OR CELL ADHESION.
 CC -1- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KDa WITH IGF-1 OR IGF-II AND IGFBP-3.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: PLASMA.
 CC -1- SIMILARITY: CONTAINS 20 LEUCINE-RICH REPEATS (LRR).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M86826; AAA36047.1; .
 CC EMBL; AL031724; CAC36078.1; .
 CC PIR; A41915; A41915.
 CC HSSP; P23945; LXUN.
 CC MIM; 601489; .
 CC InterPro; IPR001611; LRR.
 CC InterPro; IPR000483; LRR_Cterm.
 CC InterPro; IPR000372; LRR_Nterm.
 CC InterPro; IPR003592; LRR_out.
 CC InterPro; IPR003591; LRR_tyr.
 CC Pfam; PF00560; LRR_19.
 CC Pfam; PF01463; LRRCT; 1.
 CC Pfam; PF01462; LRRNT; 1.
 CC PRINTS; PR00019; LEURICHRPT.
 CC SMART; SM00370; LRR; 2.
 CC SMART; SM00882; LRRCT; 1.

```
DR SMART: SM00013; LRRNT; 1.
DR SMART: SM00369; LRR_TYP; 11.
KW Glycoprotein; Leucine-rich repeat; Repeat; Signal.
FT SIGNAL 1 27
FT CHAIN 28 605
FT REPEAT 53 73 INSULIN-LIKE GROWTH FACTOR BINDING
FT REPEAT 74 96 PROTEIN COMPLEX ACID LABILE CHAIN.
FT REPEAT 98 120 LRR 1.
FT REPEAT 121 144 LRR 2.
FT REPEAT 145 168 LRR 3.
FT REPEAT 170 192 LRR 4.
FT REPEAT 193 216 LRR 5.
FT REPEAT 217 240 LRR 6.
FT REPEAT 242 264 LRR 7.
FT REPEAT 265 288 LRR 8.
FT REPEAT 289 312 LRR 9.
FT REPEAT 313 336 LRR 10.
FT REPEAT 337 360 LRR 11.
FT REPEAT 361 384 LRR 12.
FT REPEAT 386 408 LRR 13.
FT REPEAT 409 432 LRR 14.
FT REPEAT 433 456 LRR 15.
FT REPEAT 458 480 LRR 16.
FT REPEAT 482 504 LRR 17.
FT REPEAT 505 530 LRR 18.
FT CARBOHYD 64 64 LRR 19.
FT CARBOHYD 85 85 LRR 20.
FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 580 580 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 605 AA; 66034 MW; F6562A23CB918F6 CRC64;

Query Match 10.9%; Score 407.5; DB 1; Length 605;
Best Local Similarity 25.2%; Pred. No. 6.5e-22;
Matches 152; Conservative 72; Mismatches 173; Indels 205; Gaps 16;

QY 3 LVAPELLLVNA-----GATATVVPVWVHVCPPQACQOIRPMWTPRSSY---REATT 52
DB 8 LALALLLSVVALGPRSLGADPGTGEAEGPACPAVC-----SYDDADELS 57
QY 53 VQDNDFLVAVPALPAGTQTLLOSNTYRV-----DQSEIGY----- 91
DB 58 VFSSSNRLRLPGVGGTQALMLDGNLSSVPPAFOMLSSIGFLNTLOGGOLGLEPQA 117
QY 92 ---LANLTGLDLSQNSFSDARCDFFHALPOLSLHLENQTLRIE----- 133
DB 118 LGLLENLCHLHERNQLRSLALGTFAHTPALASIGLNNRLSRLEDGLFEGIGSLMDNL 177
QY 134 -----DHSFAGIASIQEELYLNHNQYRIAPRAFSGLSNLRLHLNSNLRAIDSRW 184
DB 178 GWNLSLAVLPDAAFRGISREIYLAGNRLAYLQPALEFSLAEIRLDELISRNLRAIKANV 237
QY 185 FEMLPNLETIMIGKNVDAT-----LDMN----- 208
DB 238 FVQLPRLQTLTYDRNLILAAVACGALGLKALRWLDLNSHRVAGLLEDTFFGILGLRVRL 297
QY 209 -----FRPLANLRSYVLGAMNLRISDYALEGLQSLLESISFYDNOIARV---- 252
DB 298 SHNATASLRPTFKDLHLEELQLOHNHNRROLAERSFELQGLLEVLTIDHNOLOQYKKA 357
QY 253 -----PRALAEQVPGLEKFDLKNPNLQRVGPGDFAMNLIKELGL 292
DB 358 FLGLTNVAVMNLGNCRLNLPQVFRGLGKLSHLSESCIGRIRPHTFTGSLGRFL 417
QY 293 NNMEELVSDKPAIVNLPETLKTDTNNPRLSFTHPRAFHILPQWETMLMNNALSAL-- 350
DB 418 KD-NGLVGIEOSLWGLLELELDLTSN-QLTHTLPHRLFOGLKLEYLLLSRNRRLAEIPA 475
QY 351 -----HQQTVESLPPN-----LOEVGLHG 368
```

```
DB 476 DALGPIQRAFWLDVSHNRLEALPNSLIAPLGLRLYLISLNNSLRFTTPOPELERLMBG 535
QY 369 NPTRCDCVIR-----WANATGTRV-REIEP-----GSTLCAEPDLOQLPVR 409
DB 536 NPWDCCCPKLRDEFLQNPASAVPRVQAICGDDCCQPPATYNNITCASPREVVGDLRL 595
QY 410 EV 411
DB 596 DL 597

RESULT 3
ALS_PAPHA
ID ALS_PAPHA STANDARD; PRT; 605 AA.
AC 002833;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Insulin-like growth factor binding protein complex acid labile chain
DE precursor (ALS).
GN IGFALS OR ALS.
OS Papio hamadryas (Hamadryas baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Papio.
OX NCBI_Taxid=9557;
RX [1]
RP SEQUENCE FROM N.A.
RP TISSUE=Liver;
RC MEDLINE=97040714; PubMed=8886027;
RA Delhanty P., Baxter R.C.;
RT "The cloning and expression of the baboon acid-labile subunit of the
RT insulin-like growth factor binding protein complex.";
RL Biochem. Biophys. Res. Commun. 227:897-902(1996).
CC -! FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS THAT RESULT
CC IN PROTEIN COMPLEXES, RECEPTOR-LIGAND BINDING OR CELL ADHESION.
CC -! SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KDA WITH
CC IGF-I OR IGF-II AND IGFBP-3 (BY SIMILARITY).
CC -! SURCELLULAR LOCATION: Extracellular.
CC -! SIMILARITY: CONTAINS 20 LEUCINE-RICH REPEATS (LRR).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; S83462; -; NOT_ANNOTATED_CDS.
CC HSSP; P23945; 1XUN.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR000483; LRR_Cterm.
DR InterPro: IPR000372; LRR_Nterm.
DR InterPro: IPR003592; LRR_out.
DR InterPro: IPR003591; LRR_typ.
DR Pfam; PF00560; LRR; 19.
DR Pfam; PF01463; LRRCT; 1.
DR Pfam; PF01462; LRRNT; 1.
DR PRINTS; PRO0019; LEURICHPT.
DR SMART; SM00370; LRR; 2.
DR SMART; SM00082; LRRCT; 1.
DR SMART; SM00013; LRRNT; 1.
DR SMART; SM00369; LRR_TYP; 11.
KW Glycoprotein; Leucine-rich repeat; Repeat; Signal.
FT SIGNAL 1 27
FT CHAIN 28 605
FT REPEAT 53 73 INSULIN-LIKE GROWTH FACTOR BINDING
FT REPEAT 74 96 PROTEIN COMPLEX ACID LABILE CHAIN.
FT REPEAT 98 120 LRR 1.
FT REPEAT 121 144 LRR 2.
FT REPEAT 145 168 LRR 3.
FT REPEAT 170 192 LRR 4.
FT REPEAT 193 216 LRR 5.
```

```

FT REPEAT 169 192 LRR 6.
FT REPEAT 194 216 LRR 7.
FT REPEAT 217 240 LRR 8.
FT REPEAT 242 264 LRR 9.
FT REPEAT 265 288 LRR 10.
FT REPEAT 289 312 LRR 11.
FT REPEAT 313 336 LRR 12.
FT REPEAT 338 360 LRR 13.
FT REPEAT 361 384 LRR 14.
FT REPEAT 386 408 LRR 15.
FT REPEAT 409 432 LRR 16.
FT REPEAT 433 456 LRR 17.
FT REPEAT 458 480 LRR 18.
FT REPEAT 482 504 LRR 19.
FT REPEAT 505 530 LRR 20.
FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 580 580 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 605 AA; 66110 MW; 9D71994625F23652 CRC64;

```

```

Query Match 10.7%; Score 397.5; DB 1; Length 605;
Best Local Similarity 24.9%; Pred. No. 3,4e-21;
Matches 150; Conservative 73; Mismatches 173; Indels 207; Gaps 16;

```

```

OY 3 LVAVALLLAWA-----GATATPVVPMVHPGPCQACQIRPVYTPRSSYREAT---T 52
D 8 LALALLLSWALGPRSLGEGAEPPGEGAEACATAC-----SYDEVNELS 57
OY 53 VDCNDLFTAVAPALPAGTDTLLQSNSTIVR-----DOS 87
D 56 VCCSSNRLTRLPDIPGQALMDLSDNNLSIPPAFNLMSIAFLNQGQGLGSLPEQA 117
OY 88 ELGYLANLTEDLSONSDARDCCPHALPOLSLHLENOITLRE----- 133
D 118 LIG-LENLCHLHLENOITLSLAVGTFAYPALALIGLSNNRLSRLDEGLFEGGLMMDLN 176
OY 134 -----DHSPAGIASLOEYLNHNQYLRAPAFSGLSNLRHLNLSLRRAIDSR 183
D 177 LQWNSIAVLPDAAFRGELGRLVLAGNRLAYLQPALSSGLAELELDLSRAALAIKAN 236
OY 184 WFEMLPNEILMIGNKVDAT-----IDMN----- 208
D 237 VFAQLPRIQKIVLRNLIAAVAPGFLGLKALRWLDLSHNRVAGLLEDTFPGLGLRVLR 296
OY 209 -----FRPLANIRSLIVAGANLREISDVALLEGLOSLSLSTYDNQARV--- 252
D 297 LSHNAIASLRPTFEDLHLELELOLGHNRIRQLAERSFEGQLQLEVLTLDNHQLOEVKVG 356
OY 253 -----RRALEQVPGKFLDLPNKPORYVGPGEFANMLHLKEIG 291
D 357 AELGLTNVAVANMISGNCILRNLPQVYRGKGLSHLSLBESSCGIRPHFPAISGLRIKF 416
OY 292 LNMEEVLVIDKFAVNLDELTKLDTNNPRLSFTHRAFNHLPMQETMLNNNAISAL- 350
D 417 LKN-NGLVGIEEGSLWGLAELELDLTSN-QLTNHPHQFGIGKLELYLLSHNRLAELP 474
OY 351 -----HQQVESL-----NIDQEVGLH 367
D 475 ADAIGLQRAFWLIDVSHNRLAELPGLSLASGLRLRYLNRNNSLRTFTPPQGLELWLE 534
OY 368 GNPTRCDQVIR---WANATGTRV-RFIEP-----OSTLCAEPPDQRLP 408
D 535 GAWWDSCPLKALRDPALQNPSPAFVQALICGDDCQPPVYTYNNITGASPEVAGDL 594
OY 409 REV 411
D 595 RDL 597

```

```

RESULT 4
ALS_RAT
ID ALS_RAT STANDARD: PRT; 603 AA.
AC P35859;
DT 01-JUN-1994 (rel. 29, Created)
DT 01-JUN-1994 (rel. 29, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Insulin-like growth factor binding protein complex acid labile chain
DE precursor (ALS).
GN IGFALS OR ALS.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93038676; PubMed=1384485;
RA Dal J., Baxter R.C.;
RT "Molecular cloning of the acid-labile subunit of the rat insulin-like growth factor binding protein complex.";
RT Endocrinology 134:848-852(1994).
RL Biochem. Biophys. Res. Commun. 188:304-309(1992).
RN [2]
RP SEQUENCE OF 24-44, AND CHARACTERIZATION.
RC STRAIN=MISTAR; TISSUE=Seraum;
RX MEDLINE=94130835; PubMed=7507839;
RA Baxter R.C., Dal J.;
RT "Purification and characterization of the acid-labile subunit of rat serum insulin-like growth factor binding protein complex.";
RL Endocrinology 134:848-852(1994).
CC -1- FUNCTION: MAY HAVE AN IMPORTANT ROLE IN REGULATING THE ACCESS OF CIRCULATING IGFs TO THE TISSUES.
CC -1- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KDA WITH IGF-1 OR IGF-II AND IGFBP-3.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: BRAIN, KIDNEY, LUNG, HEART, SPLEEN, MUSCLE AND LIVER.
CC -1- SIMILARITY: CONTAINS 21 LEUCINE-RICH REPEATS (LRR).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
CC
CC EMBL; S46785; AAB23770.2; -.
CC PIR; JC1282; JC1282.
CC HSP; P23945; 1XUN.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000483; LRR_Cterm.
DR InterPro; IPR000372; LRR_Nterm.
DR InterPro; IPR003592; LRR_out.
DR InterPro; IPR003591; LRR_Typ.
DR Pfam; PF00560; LRR_19.
DR Pfam; PF01463; LRRCT; 1.
DR Pfam; PF01462; LRRNT; 1.
DR PRINTS; PR00019; LEORICHRPT.
DR SMART; SM00370; LRR; 5.
DR SMART; SM00082; LRRCT; 1.
DR SMART; SM00013; LRRNT; 1.
DR SMART; SM00369; LRR_Typ; 9.
KW Glycoprotein; Leucine-rich repeat; Repeat; signal.
FT SIGNAL 1 23
FT CHAIN 24 603
FT REPEAT 52 73 INSULIN-LIKE GROWTH FACTOR BINDING
FT REPEAT 74 96 PROTEIN COMPLEX ACID LABILE CHAIN.
FT REPEAT 98 120 LRR 1.
FT REPEAT 121 144 LRR 2.
FT REPEAT 146 168 LRR 3.
FT REPEAT 169 192 LRR 4.
FT REPEAT 192 LRR 5.
FT REPEAT 192 LRR 6.

```

```

FT REPEAT 194 216 LRR 7.
FT REPEAT 217 240 LRR 8.
FT REPEAT 242 264 LRR 9.
FT REPEAT 266 288 LRR 10.
FT REPEAT 289 312 LRR 11.
FT REPEAT 313 336 LRR 12.
FT REPEAT 337 360 LRR 13.
FT REPEAT 361 384 LRR 14.
FT REPEAT 386 408 LRR 15.
FT REPEAT 409 432 LRR 16.
FT REPEAT 433 456 LRR 17.
FT REPEAT 458 479 LRR 18.
FT REPEAT 479 504 LRR 19.
FT REPEAT 506 529 LRR 20.
FT REPEAT 543 566 LRR 21.
FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 586 586 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 603 AA: 66811 MW: DCD7637D94A5037C CRC64;

```

```

Query Match 9.2%; Score 344; DB 1; Length 603;
Best Local Similarity 23.3%; Pred. No. 2,6e-17;
Matches 141; Conservative 63; Mismatches 194; Indels 208; Gaps 16;

```

```

QY 8 LLAAWVA-----GATATVPVPMVHPVPCACQIRWYPRSSYRATVTCN 56
  |||
  |||
  |||
Db 13 LIAFWALGSGCHIQGDHPGASADA---GPPCVACTS-----HDDYDELISVPCS 61
  |||
  |||
  |||
QY 57 DLEFLAVPALPAGTQTLLOSNSIVRV-----DOSELGY 91
  |||
  |||
  |||
Db 62 SKNLTFLPDDIPVSTRALMDGNLSIPSAFQNLSDFLNLQSGWLSLEPQALLG- 120
  |||
  |||
  |||
QY 92 LANLTLELDSQNSFSDARDCDFALPQLLSLHEENQTLRL- 133
  |||
  |||
  |||
Db 121 LQNLVYLHERNLRLNLAAGLFTHPSLASLSSWILGLEGLFQGLSHLMDLNLGNW 180
  |||
  |||
  |||
QY 134 -----DHSFAGLASLOELYLNHNOYLAPRAFSGLSMLRLHLSNLRALDSWFEEM 187
  |||
  |||
  |||
Db 181 SLVLPDLPYQGLNHLHELVLNAGNKLTLYLPALFGLRELDLSRNALRSVKANFVH 240
  |||
  |||
  |||
QY 188 LPNLLELIMIGKNKVDAL-----LDMN----- 208
  |||
  |||
  |||
Db 241 LPRILQKLYLDRNLITAVAGAFGLGKALRMLDLSHNRVAGLMDPTPGLGLHYLRALHN 300
  |||
  |||
  |||
QY 209 -----FRPLANLRSLVLAGMNLREISDYALSGLSLESISFYDNLARV----- 252
  |||
  |||
  |||
Db 301 AIASLRPRFKDLHLEELQGLHNRLQROGERTFEGIQLEVLTLINDQITEVRVGAFGS 360
  |||
  |||
  |||
QY 253 -----PRRALEQVPGLEKFLDKNKPLQVRGSPGFANLHLKELGINNM 295
  |||
  |||
  |||
Db 361 LFNVAVMNLSGNCRLSLPRVROGDLKLSLHLSCHLGHVLLHTFPAGSGLRRLFLRD- 419
  |||
  |||
  |||
QY 296 EEELVSDKALVNLPELTLDITNNPRLSFTHPRAFHNLPMQETMLNNALSAV----- 350
  |||
  |||
  |||
Db 420 NSISSIEQSLAGLSLELDTLTN-RLTHLPRQLFQGLGHEVLLLSVNOJLTLSAEVL 478
  |||
  |||
  |||
QY 351 -----HQ-----QVESLPMIQEVLGHGNFIR 372
  |||
  |||
  |||
Db 479 GELQRAFWLIDISHNLLETLAGLSLGRVRYLSLRNNSLQLFSPQGLERLMDLANPMD 538
  |||
  |||
  |||
QY 373 CQCVIR-----WANATGRVRFIEP-----QSTLCAPPPDLQRLPVREVPFR 414
  |||
  |||
  |||
Db 539 CSCPLKALBDALQNPVGVPFRVQTVCSGSDQCPVYITNNITCAGPNAVSGLDLKVSET 598
  |||
  |||
  |||
QY 415 EMTDHC 420
  |||
  |||
  |||
Db 599 HRFV-HC 603

```

```

RESULT 5
ALS_MOUSE STANDARD; PRT; 603 AA.
ID ALS_MOUSE
AC P70389;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Insulin-like growth factor binding protein complex acid labile chain
DE precursor (ALS).
DE IGFBP3 OR ALS OR ALBS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OX Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.
RN NCBI_Taxid=10090;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=96413591; PubMed=8816745;
RA Boisclair Y.R., Seto D., Hsieh S., Hurst K.R., Ooi C.T.;
RT "Organization and chromosomal localization of the gene encoding the
RT mouse acid labile subunit of the insulin-like growth factor binding
RT complex."
RL Proc. Natl. Acad. Sci. U.S.A. 93:10028-10033(1996).
CC -!- FUNCTION: MAY HAVE AN IMPORTANT ROLE IN REGULATING THE ACCESS OF
CC CIRCULATING IGFs TO THE TISSUES.
CC -!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KDA WITH
CC IGF-1 OR IGF-TI AND IGFBP-3 (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 21 LEUCINE-RICH REPEATS (LRR).
CC -!- This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U66900; AB017270.1; -.
CC MGD: MGI:107973; Tgfbals.
CC InterPro: IPR001611; LRR.
CC InterPro: IPR000483; LRR_Cterm.
CC InterPro: IPR000372; LRR_Nterm.
CC InterPro: IPR003592; LRR_out.
CC InterPro: IPR003591; LRR_cyp.
CC Pfam: PF00560; LRR_19.
CC Pfam: PF01463; LRRCT; 1.
CC Pfam: PF01462; LRRNT; 1.
CC PRINTS: PR00019; LEURICHRPT.
CC SMART: SM00370; LRR; 4.
CC SMART: SM00082; LRRCT; 1.
CC SMART: SM00013; LRRNT; 1.
CC SMART: SM00369; LRR_Typ; 9.
CC Glycoprotein; Leucine-rich repeat; Repeat; Signal.
CC SIGNAL 1 23
CC CHAIN 24 603
FT FT 52 73 LRR 1.
FT FT 96 96 LRR 2.
FT FT 120 120 LRR 3.
FT FT 121 144 LRR 4.
FT FT 146 168 LRR 5.
FT FT 169 192 LRR 6.
FT FT 194 216 LRR 7.
FT FT 217 240 LRR 8.
FT FT 242 264 LRR 9.
FT FT 265 288 LRR 10.
FT FT 289 312 LRR 11.
FT FT 313 336 LRR 12.
FT FT 338 360 LRR 13.
FT FT 361 384 LRR 14.
FT FT 386 408 LRR 15.

```

FT REPEAT 409 432 LRR 16.
 FT REPEAT 433 456 LRR 17.
 FT REPEAT 458 480 LRR 18.
 FT REPEAT 482 504 LRR 19.
 FT REPEAT 505 529 LRR 20.
 FT REPEAT 543 566 LRR 21.
 FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 586 586 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 603 AA; 66959 MW; 4A6223AEC7026BCC CRC64;

Query Match 9.2%; Score 342.5; DB 1; Length 603;
 Best Local Similarity 22.0%; Pred. No. 3,3e-17;
 Matches 139; Conservative 69; Mismatches 190; Indels 233; Gaps 15;

8 LLLAWA-----GATATVPPVWVWVPCPOCAQIRPWTPSSSTREATTVVCN 56
 13 LLAFWALGPCYLGCTDGPASADAE---GPOCAPVTCTCSY-----DDYTDLSVPCS 61
 57 DLFIRAVPALPAGNOTILLGNSIVRY-----DOSELGX 91
 62 SRNLQPLPGITVSTRALMDGNNSLSPSAFQNLSSDLFNLOGSMRLSEFQALLG- 120
 92 LANLELDLSONSFSDARCDFFHALPQILLHLLENOITRLE----- 133
 121 LQNLVHLHERNLRLSLAGLFRHTPPLASLSLGNMLGRLEEGFRGLSHLMDLGN 180
 134 -----DHFAGLASIQEILYLNHNQYIARAPAFSGLSLRLHLNSLRLAIDSRWFEM 187
 181 SLUVLPDTPVFOGNGNLHLVYLGKNTLYLQPALGLGELRELDLSRNALNSVKANVEIH 240
 188 LPLLEIMIGKNKVDAL-----LDMN----- 208
 241 LPRLOKLYIDRNLTAVARAFGLKMKALKWLDSHNRAVGLEDIEFPGLLGHVRLAHN 300
 209 -----FRPLANIRSLVLAGMNIRESIDVALBSGLSELSFYDNLAV----- 252
 301 AITSLRPRTFKDLHLEELQLOLGNHNRIRQGEKTFEGLGQLEVLTTDNDQIHVEKVGAFEG 360
 253 -----PRALAEQVPGKFLDKNKPLIRVBPGRFANMLHKEGLINNM 295
 361 LENVAVMNISGNCRLSLRPEHVFQGLRLSHLHESCLRIKHTFAGLSGLRLFLRD- 419
 296 EELVSTDKFALVNIPELTIKLIDITNNRSLFIHPRAFHLPOMETLMLNNALSAHQ--- 352
 420 NSISSTIEGSLAGLSLELDELDTAN-QLTHLPRLQFLGQLEVLLSNNQITMLSEVYL 478
 353 -----QYVESLPNIOEVLGNPNIR 372
 479 GPLORAFWLDSLNRLETPAEGFLSSGLRLRYLNLNRNSLQTFVPPQGLERLMLDANPMD 538
 373 CDCVIR-----WANATGTRVRFIEP-----OSTICAEPPDDQLRPVREVPFR 414
 539 CSCPLKALRFALONPGVVRFYQVCEGDQCPVYTYNNITCAGANYSGLDLRII--- 595
 415 EMTDHLCLPLISPRSPPSLOVASGESMVLHC 445
 596 -----SETLFLVHC 603

GN GP5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Liver;
 RX MEDLINE=97275136; PubMed=9129030;
 RA Ravanat C., Morales M., Azotza D.O., Moog S., Schubler S.,
 RA Guntert P., Loew D., van Dorsselaer A., Cazenave J.-P., Lanza F.,
 RT "Gene cloning of rat and mouse platelet glycoprotein V:
 RT Identification of megakaryocyte-specific promoters and demonstration
 RT of functional thrombin cleavage.";
 RL Blood 89:3253-3262(1997).
 CC - FUNCTION: THE GPIIb-IIIx COMPLEX FUNCTIONS AS THE VON WILLEBRAND
 CC FACTOR RECEPTOR AND MEDIATES VON WILLEBRAND FACTOR-DEPENDENT
 CC PLATELET ADHESION TO BLOOD VESSELS. THE ADHESION OF PLATELETS TO
 CC INJURED VASCULAR SURFACES IN THE ARTERIAL CIRCULATION IS A
 CC CRITICAL INITIATING EVENT IN HEMOSTASIS (BY SIMILARITY).
 CC - SUBCELLULAR LOCATION: Type I membrane protein.
 CC - SIMILARITY: CONTAINS 14 LEUCINE-RICH REPEATS (LRR).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC EMBL; Z69595; CAA93441.1; -
 DR MGD; MGI:1096363; GP5.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR000483; LRR_Cterm.
 DR InterPro: IPR003372; LRR_Nterm.
 DR InterPro: IPR003592; LRR_Out.
 DR InterPro: IPR003591; LRR_Cyp.
 DR Pfam; PF00560; LRR; 13.
 DR Pfam; PF01463; LRRCT; 1.
 DR PRINTS; PR00019; LEURICHRT.
 DR SMART; SM00370; LRR; 1.
 DR SMART; SM00082; LRRCT; 1.
 DR SMART; SM00013; LRRNT; 1.
 DR SMART; SM00369; LRR_Typ; 10.
 DR Platelet; Transmembrane; Glycoprotein; Blood coagulation;
 KW Repeat; Leucine-rich repeat; Cell adhesion; Signal.
 FT SIGNAL 1 16
 FT CHAIN 17 567
 FT DOMAIN 17 522
 FT TRANSMEM 523 543
 FT DOMAIN 544 567
 FT REPEAT 73 96
 FT REPEAT 97 120
 FT REPEAT 122 144
 FT REPEAT 145 168
 FT REPEAT 170 192
 FT REPEAT 194 216
 FT REPEAT 217 240
 FT REPEAT 241 264
 FT REPEAT 266 288
 FT REPEAT 289 312
 FT REPEAT 314 337
 FT REPEAT 338 361
 FT REPEAT 362 385
 FT REPEAT 386 409
 FT CARBOHYD 51 51
 FT CARBOHYD 67 67
 FT CARBOHYD 181 181
 FT CARBOHYD 243 243
 FT CARBOHYD 298 298
 FT CARBOHYD 312 312
 FT CARBOHYD 385 385

FT	REPEAT	765	788	LRR 21.
FT	REPEAT	790	812	LRR 22.
FT	REPEAT	813	836	LRR 23.
FT	REPEAT	838	861	LRR 24.
FT	DOMAIN	907	944	EGF-LIKE 1.
FT	DOMAIN	946	983	EGF-LIKE 2.
FT	DOMAIN	985	1022	EGF-LIKE 3.
FT	DOMAIN	1024	1062	EGF-LIKE 4.
FT	DOMAIN	1064	1100	EGF-LIKE 5.
FT	DOMAIN	1111	1149	EGF-LIKE 6.
FT	DOMAIN	1152	1325	EGF-LIKE 7.
FT	DOMAIN	1353	1392	EGF-LIKE 7.
FT	DOMAIN	1409	1480	CRCK.
FT	CARBOHYD	111	111	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	207	207	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	357	357	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	435	435	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	783	783	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	788	788	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	958	958	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	998	998	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1060	1060	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1159	1159	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1175	1175	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1243	1243	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1292	1292	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	DISULFID	911	922	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	DISULFID	916	932	BY SIMILARITY.
FT	DISULFID	934	943	BY SIMILARITY.
FT	DISULFID	950	961	BY SIMILARITY.
FT	DISULFID	955	971	BY SIMILARITY.
FT	DISULFID	973	982	BY SIMILARITY.
FT	DISULFID	989	1001	BY SIMILARITY.
FT	DISULFID	995	1010	BY SIMILARITY.
FT	DISULFID	1012	1021	BY SIMILARITY.
FT	DISULFID	1028	1041	BY SIMILARITY.
FT	DISULFID	1035	1050	BY SIMILARITY.
FT	DISULFID	1052	1061	BY SIMILARITY.
FT	DISULFID	1068	1079	BY SIMILARITY.
FT	DISULFID	1073	1088	BY SIMILARITY.
FT	DISULFID	1090	1099	BY SIMILARITY.
FT	DISULFID	1115	1125	BY SIMILARITY.
FT	DISULFID	1120	1137	BY SIMILARITY.
FT	DISULFID	1139	1148	BY SIMILARITY.
FT	DISULFID	1357	1368	BY SIMILARITY.
FT	DISULFID	1362	1380	BY SIMILARITY.
FT	DISULFID	1382	1391	BY SIMILARITY.
FT	DISULFID	1409	1443	BY SIMILARITY.
FT	DISULFID	1423	1457	BY SIMILARITY.
FT	DISULFID	1434	1473	BY SIMILARITY.
FT	DISULFID	1438	1475	BY SIMILARITY.
FT	DISULFID	1442	1479	BY SIMILARITY.
FT	DISULFID	1394	1404	BY SIMILARITY.
FT	VANSPPLIC	1480	165752	MISSING (IN SHORT ISOFORM).
SO	SEQUENCE	AA:	MM;P905925FC170D1C3	CRG64;

Query Match	8.7%	Score 324	DB 1	Length 1480
Best Local Similarity	24.7%	Pred. No. 2.6e-15		
Matches 103	Conservative	67	Mismatches 169	Indels 78
				Gaps 11
QY	29	CPPOCAOQIRPYPTPRSSREATYVDCNDLFLTAAPPALPACTQTLLLOSNISIVERDOSE	88	
DB	519	CPAMCHC-----ESTYVDCGRRLKEIETPROPIPTLTTELLDNLDELGRSSDG	565	
QY	89	IGYLANLTETLDSQNSFSDARDCEFHAPOLLSTLHEENQLTRELDHSEFAGIASLOEVL	148	
DB	566	L-----FGRLPHTLYKLETKRQNLGTGIEBNAEFGASHIQEQL	602	
QY	149	NINQLTIRAPRAFSGLSNLRHLNSNLRALDSKRFEMLPWLEIMIGKNVADALDMN	208	
DB	603	GENKIKESNNKFLGJHQTKTLNLVDNQISCVPGSFEHLNLSLTSLNASNFCNC-NCH	660	
QY	209	FEPILAN-IRSLIVLAGMNLK-----ETSDPALEGLOSLF-----	240	

Db	661	LAMFAECYKRSKSLNGAARCAPSKVDYOIKDLPHSEFFKCSSENSFGCLGDCYCPSPCN	720
Qy	241	-----SLSTPYDQGLARVRRRALEQYPG-LKRLDNLKNPGLRVGGDDPAMMLHLKELGLNNM	295
Db	721	CTGVVACASRNQLEKIPR---GIPATSELYESNETEDQIHYERIKHMLSLRLDLSN	775
Qy	296	EELVISIDEFALVYNPELTKLTDITNNPSPHFIRPAFHHLPMOETMLNNALSAHQQT	355
Db	776	NQITLSVYTFANLTKLSTLITSVN-KLQGLQRLALSGLNNLRVSLSHGRISIMLDEGSP	834
Qy	356	ESLPNLQEVGLHGNPIRCDVCYIRANATGTVRRFIEQOSTLCAEPPLQDLPVREVP	412
Db	835	EDLSLTFHIALGSLNPLYCDGLKWF-SDWIKLDVYEPGILARCAEPQMDKLITLSTP	890
RESULT 8			
GPY_HUMAN	8	GPY_HUMAN	STANDARD; PRT; 560 AA.
AC	P40197;		
DT	01-FEB-1995 (Rel. 31, Created)		
DT	01-FEB-1995 (Rel. 31, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Platelet glycoprotein V precursor (GPV) (CD42D).		
GN	GP5.		
OC	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
OX	NCBI_Taxid=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Lung;		
RX	MEDLINE=93391348; PubMed=7690959;		
RA	Hickey M.J., Hagen F.S., Yagil M., Roth G.J.;		
RT	"Human platelet glycoprotein V: characterization of the polypeptide		
RT	and the related Ib-V-IX receptor system of adhesive, leucine-rich		
RT	glycoproteins.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 90:8327-8331(1993).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Platelet;		
RX	MEDLINE=94012616; PubMed=8407908;		
RA	Lanza F., Morales M., de la Salle C., Cazenave J.-P., Clemetson K.J.;		
RA	Shimomura T., Phillips D.R.;		
RT	"Cloning and characterization of the gene encoding the human platelet		
RT	glycoprotein V. A member of the leucine-rich glycoprotein family		
RT	cleaved during thrombin-induced platelet activation.";		
RL	J. Biol. Chem. 268:20801-20807(1993).		
RN	[3]		
RP	TISSUE=Platelet;		
RX	MEDLINE=90275263; PubMed=2350580;		
RA	Shimomura T., Fujimura K., Maehama S., Takemoto M., Oda K.,		
RA	Fujimoto Y., Ojama R., Suzuki M., Ichihara-Tanaka K., Tlilani K.,		
RA	Kuramoto A.;		
RT	"Rapid purification and characterization of human platelet		
RT	glycoprotein V: the amino acid sequence contains leucine-rich		
RT	repetitive modules as in glycoprotein Ib.";		
RL	Blood 75:2349-2356(1990).		
RN	[4]		
RP	PARTIAL SEQUENCE.		
RC	TISSUE=Platelet;		
RX	MEDLINE=90321220; PubMed=2372284;		
RA	Roth G.J., Church T.A., McMullen B.A., Williams S.A.;		
RT	"Human platelet glycoprotein V: a surface leucine-rich glycoprotein		
RT	related to adhesion.";		
RL	Biochem. Biophys. Res. Commun. 170:153-161(1990).		
CC	-I- FUNCTION: THE GPIB-V-IX COMPLEX FUNCTIONS AS THE VON WILLEBRAND		
CC	FACTOR RECEPTOR AND MEDIATES VON WILLEBRAND FACTOR-DEPENDENT		
CC	PLATELET ADHESION TO BLOOD VESSELS. THE ADHESION OF PLATELETS TO		
CC	INURED VASCULAR SURFACES IN THE ARTERIAL CIRCULATION IS A		
CC	CRITICAL INITIATING EVENT IN HEMOSTASIS.		
CC	-I- SUBCELLULAR LOCATION: Type I membrane protein.		

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/cc> or send an email to license@isb-sib.ch).

CC CC -1- TISSUE SPECIFICITY: PLATELETS AND MEGAKARYOCYTES.
CC CC -1- PTM: THE N-TERMINUS IS BLOCKED.
CC CC -1- SIMILARITY: CONTAINS 14 LEOCINE-RICH REPEATS (LRR).

DR DR EMBL; L11238; AAA03069.1; -.
DR DR HSSP; P16473; IXUM.
DR DR MIM; I73511; -.
DR DR InterPro; IPR001611; LRR.
DR DR InterPro; IPR0000483; LRR_Cterm.
DR DR InterPro; IPR000372; LRR_Nterm.
DR DR InterPro; IPR003592; LRR_Out.
DR DR InterPro; IPR003591; LRR_Typ.
DR DR Pfam; PF00560; LRR; 14.
DR DR Pfam; PF01463; LRRC; 1.
DR DR PRINTS; PRO0019; LEURICHRPT.
DR DR SMART; SMO0370; LRR; 1.
DR DR SMART; SMO0082; LRRC; 1.
DR DR SMART; SMO0013; LRRT; 1.
DR DR SMART; SMO0369; LRR_Typ; 9.
KW Platelet; Transmembrane; Glycoprotein; Blood coagulation;
KW Repeat; Leucine-rich repeat; Cell adhesion; Signal.
FT FT SIGNAL 1 16
FT FT CHAIN 17 560
FT FT DOMAIN 17 523
FT FT TRANSHEM 524 544
FT FT DOMAIN 545 560
FT FT REPEAT 73 96
FT FT REPEAT 97 120
FT FT REPEAT 122 144
FT FT REPEAT 145 168
FT FT REPEAT 169 192
FT FT REPEAT 194 216
FT FT REPEAT 217 240
FT FT REPEAT 241 264
FT FT REPEAT 266 288
FT FT REPEAT 289 312
FT FT REPEAT 314 337
FT FT REPEAT 338 361
FT FT REPEAT 362 385
FT FT REPEAT 386 409
FT FT CARBOHYD 51 51
FT FT CARBOHYD 181 181
FT FT CARBOHYD 243 243
FT FT CARBOHYD 267 267
FT FT CARBOHYD 298 298
FT FT CARBOHYD 312 312
FT FT CARBOHYD 385 385
FT FT CARBOHYD 499 499
FT FT CONFLICT 73 74
FT FT CONFLICT 109 109
FT FT CONFLICT 130 130
FT FT CONFLICT 136 138
FT FT CONFLICT 209 209
FT FT CONFLICT 267 267
FT FT CONFLICT 327 327
FT FT CONFLICT 478 478
FT FT CONFLICT 509 509
SQ SEQUENCE 560 AA; 60959 MW; B1CB04AF8AF7J15 CRC64;

Query Match 8.6%; Score 321; DB 1; Length 560;
Best Local Similarity 26.6%; Pred. No. 1.1e-15;
Matches 141; Conservative 57; Mismatches 227; Indels 106; Gaps 16

QY	9	ILAMVAGATATVYVYVWPHNCPPOCAQGLRPWYTPRSSYREKATYDQNDLF--LTAVPRAL	67
QY	9	ILAMVAGATATVYVYVWPHNCPPOCAQGLRPWYTPRSSYREKATYDQNDLF--LTAVPRAL	67
Db	6	LICAVIGLLRAQD-----FPCPPACK-----VFDDAQSGGIVARISAL--GL	48
QY	68	PAG-TQTLILGNSIVYVDQSEIGYLVANITELDLSQNSFSDARDCDFNALPOLYHLEE	12
Db	49	PTMLITLILFGMGKRGVLOQSGFSG--MTVLQRLMISDSHISAAPGFSOLLIKTKLRLSR	10
QY	127	NQLTRDEHSFAGLASQLELYLNHNOQVRIAPRASGSLNLRHLNLSMLRAIDSMPF	18
Db	108	NKITHLPGALDKMVLLEOLFEDHNAIRGIDIDNMOKVLVNIQLEALMNQIDFLPASLFT	16
QY	187	MLPNIETLIGGNKVDAILIDMNFPRLANRSYVLGAMNIREISDYALEGISTESTPYD	24
Db	168	NLENKLIDLSGNNLTHLPKGLLGQAOKIERLLHSNRLVSLDSGLNSLGAATLQHR	22
QY	247	NOLARYPRALQVPELQKFLDNK-----NPLQRVGPGDFA	28
Db	228	NHRSIAPGAFRPNLSLSTLSRNLHAFELPSALFHSNLTLLTFEENPLAEPLGVLF	28
QY	283	NMLHKEIGELNNMEELVSDKFAVNLPELTKIDTNPRLSFIRAFHNLQWETML	34
Db	288	ENGGLQELMLNR-TQLRTLPALAFNLSLRILGVTLSPUSALPQAGAFQGGELQVIAL	34
QY	343	NNNALSAI-----HQD-----T	35
Db	347	HSNGGTALPDGLRGLGKLRQVSLRRNRRLRALPRALFNLSLSTSVQDNDHQLETLPDGY	40
QY	355	VESLNLDEVGNGHNPICDCC--VIRMANATGRVRFIEPQSTLCAEPDQLRPLRE	41
Db	407	FCALPRLEVLGLHNSWRCDCGLPFCLMLRHLGLVGESEPR--CAGPAHAGLPLMA	46
QY	411	VPRFMTDHCPLISPRSPSPSLQVASESMWLCRALAEPEPEYI--WTP	460
Db	465	LPQDA--EC---PGRQGPVPPRADSSSEAPVH--PALAPSSSEPWMAQ	509
RESULT	9		
TLR3_MOUSE		STANDARD;	PRT; 905 AA.
AC	099MB1; G912M4;		
DT	01-MAR-2002 (Rel. 41, Created)		
DT	01-MAR-2002 (Rel. 41, Last sequence update)		
DT	01-MAR-2002 (Rel. 41, Last annotation update)		
DE	Toll-like receptor 3 precursor.		
GN	TLR3.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxId=10090;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=129/Sv;		
RX	pubmed-11607032;		
RA	Alxopoulou L., Holt A.C., Medzhitov R., Flavell R.A.;		
RT	"Recognition of double-stranded RNA and activation of NF-kappaB by		
RL	Toll-like receptor 3".		
RL	Nature 413:732-738(2001).		
CC	[2]		
CC	SEQUENCE FROM N.A.		
CC	STRAIN=BALB/c x NIH; TISSUE=Macrophage;		
CC	Appelquist S.E., Ljunggren H.G.;		
CC	"Molecular cloning of mouse Toll-like receptor 3 cDNA.";		
CC	Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.		
CC	-1- FUNCTION: Participates in the innate immune response to microbial		
CC	agents. May be involved in the recognition of ds-RNA. Acts via		
CC	MyD88 and TRAF6, leading to NF-kappa-B activation, cytokine		
CC	secretion and the inflammatory response (By similarity).		
CC	-1- SUBUNIT: Binds Myd88 via their respective TIR domains (By		
CC	similarity).		
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).		
CC	-1- TISSUE SPECIFICITY: Highly expressed in lung. After		
CC	interperitoneal injection of lipopolysaccharide, highly expressed		

CC in brain, heart, kidney, liver, lung and spleen.
 CC -1- SIMILARITY: BELONGS TO THE TOLL-LIKE RECEPTOR FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 TIR DOMAIN.
 CC -1- SIMILARITY: CONTAINS 22 LEUCINE-RICH REPEATS (LRR).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF35152; AAK26117.1; -;
 DR EMBL: AF35152; AAK26117.1; -;
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR000483; LRR_Cterm.
 DR InterPro: IPR000372; LRR_Nterm.
 DR InterPro: IPR003592; LRR_out.
 DR InterPro: IPR003591; LRR_typ.
 DR InterPro: IPR000157; TIR.
 DR Pfam: PF00560; LRR; 17.
 DR Pfam: PF01463; LRRCT; 1.
 DR Pfam: PF01582; TIR; 1.
 DR PRINTS: PR00019; LEURICHRPT.
 DR SMART: SM00082; LRRCT; 1.
 DR SMART: SM00013; LRRNT; 1.
 DR SMART: SM00369; LRR_typ; 16.
 DR SMART: SM00255; TIR; 1.
 DR PROSITE: PSS0104; TIR; 1.
 KW Receptor; Immune response; Inflammatory response; Signal;
 KW Transmembrane; Repeat; Leucine-rich repeat; Glycoprotein.
 FT SIGNAL 1 25
 FT CHAIN 1 25
 FT DOMAIN 26 705
 FT TRANSMEM 706 726
 FT DOMAIN 727 905
 FT REPEAT 51 74
 FT REPEAT 76 98
 FT REPEAT 99 122
 FT REPEAT 124 146
 FT REPEAT 148 170
 FT REPEAT 171 194
 FT REPEAT 197 220
 FT REPEAT 248 271
 FT REPEAT 274 297
 FT REPEAT 298 321
 FT REPEAT 321 355
 FT REPEAT 355 380
 FT REPEAT 380 404
 FT REPEAT 404 432
 FT REPEAT 432 455
 FT REPEAT 456 479
 FT REPEAT 481 505
 FT REPEAT 506 529
 FT REPEAT 529 553
 FT REPEAT 553 585
 FT REPEAT 585 609
 FT REPEAT 609 633
 FT REPEAT 633 660
 FT REPEAT 660 697
 FT DOMAIN 697 755
 FT CARBOHYD 53 53
 FT CARBOHYD 58 58
 FT CARBOHYD 71 71
 FT CARBOHYD 125 125
 FT CARBOHYD 197 197
 FT CARBOHYD 248 248
 FT CARBOHYD 253 253
 FT CARBOHYD 276 276
 FT CARBOHYD 292 292
 FT CARBOHYD 399 399
 FT CARBOHYD 414 414
 FT CARBOHYD 425 425

FT CARBOHYD 508 508 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 663 663 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 670 670 S -> F (IN REF. 2).
 SQ SEQUENCE 905 AA; 103670 MW; 8E6DBDA9818E14B4 CMC64;
 Query Match 8.4%; Score 314; DB 1; Length 905;
 Best Local Similarity 23.4%; Pred. No. 6,8e-15;
 Matches 118; Conservative 58; Mismatches 151; Indels 178; Gaps 14;
 QY 36 QIRWYTPRSSYREATTVDNCDFITAVPAPLACAGTQTLTLOSNSIVRQDSRLGYL-A 93
 Db 233 QLNPHLTKR-----CWEIISNTSI-----QNSLANNQILANSESTFSGIKMT 275
 QY 94 NLTELDSONSFDPARCDCHALPOLSLHLEENQTLREDEHFSAGIASLOEL----- 146
 Db 276 NLTDLDLSTYNLHDVNGSFYSYLSRYLSLEVNNTORLSRPSFYGSLNRYLSLAKRAF 335
 QY 147 -----YLN----- 149
 Db 336 KQSVSLASHPNIDDFSPQMLKYLEYLIMDDNNIPSTKSNFTGLVSLKYLSTKFTSLQ 395
 QY 150 -----HNOYRIAPRAS-----G 163
 Db 396 TLTEFEVSLASHPELTTLNTRKNIISKIANGTSWGLQLILDLGLNETRQKISGQEWK 455
 QY 164 LSNLLRLHNSNLRLAIDSRWFEMLPNLEIIM--IGKNKVDAILDMFRPLANLSLYL 220
 Db 456 LRFTEIYLSNKLQLOLSTSSFALVPSLOQLMRVALKNVD-ISPSPFLPNLTLIDL 514
 QY 221 AGMNLREISDYALBGLQSLSLSFYDNOQLARVPRA-----LEQVPLKFLDLKNP 272
 Db 515 SNNNIANIEMDLBGLNLETLIDFQHNHNLARLWKRANPGPVNFKGLSHLITLINESG 574
 QY 273 LQRGPGDFAMKLK--ELGINMEEVLSIDKFAVLNLEIKLDTNNPRLSFIHRA 330
 Db 575 IDEIPVGFKNLFEKLSINGLNMLNKL-----EPTI 606
 QY 331 FHHLPQMETLMLNNNALSAHLHQTV-ESLPNLQEVGLHGNPIRDC-----VIKMANATG 384
 Db 607 FDDQTSLSRLNLQKNLITSVKDFGPPFQNLNSLDKRFNPDCTCSISWFWINQTH 666
 QY 385 TRVREIPEQSTLCAEPDQLRPLVR 409
 Db 667 TNISELSTH-YLCNTPHHYGFPLK 690
 RESULT 10
 GPV_RAT STANDARD: PRT: 567 AA.
 ID GPV_RAT
 AC 008770;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Platelet glycoprotein V precursor (GPV) (CD42b).
 GN GPV.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=MISTAR; TISSUE=Liver;
 RA MEDLINE=97275136; PubMed=9129030;
 RA Ravanat C., Morales M., Azora D.O., Moog S., Schuhler S.,
 RA Grunert P., Loew D., van Dorsselaer A., Gazeauve J.-P., Lanza F.;
 RT "Gene cloning of rat and mouse platelet glycoprotein V:
 RT identification of megakaryocyte-specific promoters and demonstration
 RT of functional thrombin cleavage.";
 RT Blood 89:3253-3262(1997).
 RL Blood 89:3253-3262(1997).
 CC -1- FUNCTION: THE GPIIb-V-III COMPLEX FUNCTIONS AS THE VON WILLEBRAND
 CC FACTOR RECEPTOR AND MEDIATES VON WILLEBRAND FACTOR-DEPENDENT
 CC PLATELET ADHESION TO BLOOD VESSELS. THE ADHESION OF PLATELETS TO

INURED VASCULAR SURFACES IN THE ARTERIAL CIRCULATION IS A
 CRITICAL INITIATING EVENT IN HEMOSTASIS (BY SIMILARITY).
 - SUBCELLULAR LOCATION: Type I membrane protein.
 - SIMILARITY: CONTAINS 14 LEUCINE-RICH REPEATS (LRR).

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 CC EMBL: Z69594; CAA93440.1; -
 CC InterPro: IPR001611; LRR.
 CC InterPro: IPR000483; LRR_Cterm.
 CC InterPro: IPR000372; LRR_Nterm.
 CC InterPro: IPR003592; LRR_out.
 CC Pfam: PF00560; LRR; 14.
 CC Pfam: PF01463; LRRCT; 1.
 CC PRINTS: PR00019; LEURICHRPT.
 CC SMART: SM00082; LRR; 2.
 CC SMART: SM00082; LRRCT; 1.
 CC SMART: SM00013; LRRNT; 1.
 CC SMART: SM00369; LRR_Typ; 10.
 CC Platelet; Transmembrane; Glycoprotein; Blood coagulation;
 KW Repeat; Leucine-rich repeat; Cell adhesion; Signal.
 FT SIGNAL 1 16 POTENTIAL.
 FT CHAIN 17 567 PLATELET GLYCOPROTEIN V.
 FT DOMAIN 17 522 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 523 543 POTENTIAL.
 FT DOMAIN 544 567 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 73 96 LRR 1.
 FT REPEAT 97 120 LRR 2.
 FT REPEAT 122 144 LRR 3.
 FT REPEAT 145 168 LRR 4.
 FT REPEAT 169 192 LRR 5.
 FT REPEAT 194 216 LRR 6.
 FT REPEAT 217 240 LRR 7.
 FT REPEAT 241 264 LRR 8.
 FT REPEAT 266 288 LRR 9.
 FT REPEAT 289 312 LRR 10.
 FT REPEAT 314 337 LRR 11.
 FT REPEAT 338 361 LRR 12.
 FT REPEAT 362 385 LRR 13.
 FT REPEAT 387 409 LRR 14.
 FT CARBOHYD 51 51 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 243 243 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 312 312 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 385 385 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 567 AA; 63344 MW; CA10708ED03707F CRC64;

 Query Match 8.4%; Score 313.5; DB 1; Length 567;
 Best Local Similarity 24.0%; Pred. No. 3.8e-15;
 Matches 133; Conservative 62; Mismatches 226; Indels 133; Gaps 13;

 QY 4 LVAPLLAMVAGATATVYVPMHVPCCPOGACQIRPWTPSSREATVVCNLFLETAV 63
 DB 1 MURSVLSAVLSVGAQP-----PPCPRTCKCVKR-----DAVCCSGGSVAHI 43
 QY 64 PP-ALPAG-----TQTLLOSNSIYVVDQSELGYANTLET 98
 DB 44 AELGLPTMLTILLFRMDRGVLQSHSFGMTVLRIMLSDSHIAIDPGTFNDLVKLT 103
 QY 99 DLQSQSFSDADCDGFHALPOLSLHLEBNQTLREDSHFGSLASLOELYNHNOLYRIAP 158
 DB 104 RLTRNKISHPRAILDKWVLLLEOLFDDHNAFLDDQNLFOKLNLRLCLINQNLSPFLPA 163

QY 159 RAESGLSNLRHLNLSNLLRAIDSRWFEMLPNLEIMIGCKKVDAILDMNPRLANRSL 218
 DB 154 NLFSGLGKLKVDLSRNNLTHLPQGLLAQTKLEKLLYSNR--LMSLDSGLLANGAL 220
 QY 219 V---LAGNLRREISDYALEGLQSLSEISFYDNOIARVPRRALEQVPKLEFDLKNPJOR 275
 DB 221 TELRLERHNRHSIAPGAFDSLGNTSTLISGNLLESIPRALFLHVSMTLTLTFENLEE 280
 QY 276 VGPQGFAMMLHLKELGNNMEELVSIDKFAIVNPDELTKDITNNPRLSTHPRAFHNP 335
 DB 281 LPEVLFGEMAGLRRLWNG--THLRTLPAAAFRNISGQTLGLTRNPLLSALPGMFGHGL 339
 QY 336 QMETLMLNNALSAHQVTESLPNLOVGLHGNPR----- 372
 DB 340 ELRLVAVHTNLEELPEALRGRLQVSLRHRRLALPRTLEFRNLSSLVTVQLEHNO 399
 QY 373 -----CDC---VIRMANATGRVRFTEPOSTLCAEPD 403
 DB 400 KTLPGDVFAALPOLTRVLLGHNPWLCDCGLMPFLQWLRHHELLGRDEPQ--CNGPESR 457
 QY 404 QRLPV-----REVPRFEMTDHCLPLISPSFPPSLQ-----VASGESMVL 443
 DB 458 ASLTFWELLQGDQWCPSSRGLPDPPTENALKAPDPYQRPNSSQSMWQVLVARGES--- 514
 QY 444 HCRALAEPEPEIYW 457
 DB 515 -----PDNRFYW 521

 RESULT 11
 ID CHAD_BOVIN STANDARD; PRT; 361 AA.
 AC Q27972;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Chondroaderlin precursor (Cartilage leucine-rich protein) (38 kDa bone
 DE protein).
 GN CHAD.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Cartilage;
 RC MEDLINE=94342341; PubMed=8063792;
 RA Neame P.J., Sommarin Y., Boynton R.E., Helnegard D.;
 RT "The structure of a 38-kDa leucine-rich protein (chondroaderlin)
 RT isolated from bovine cartilage."
 RL J. Biol. Chem. 269:21547-21554(1994).
 RN [2]
 RP SEQUENCE OF 25-55 AND 77-97.
 RC TISSUE=Bone;
 RC MEDLINE=95113864; PubMed=7814406;
 RA Hu B., Coulson L., Moyer B., Price P.A.;
 RT "Isolation and molecular cloning of a novel bone phosphoprotein
 RT related in sequence to the cysstatin family of thiol protease
 RT inhibitors."
 RL J. Biol. Chem. 270:431-436(1995).

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 CC EMBL: U08018; AAA21330.1; -
 CC InterPro: IPR001611; LRR.
 CC InterPro: IPR000483; LRR_Cterm.

```

DR InterPro; IPR000372; LRR_Nterm.
DR InterPro; IPR003591; LRR_Typ.
DR Pfam; PF00560; LRR; 10.
DR Pfam; PF01463; LRRCT; 1.
DR Pfam; PF01462; LRRNT; 1.
DR SMART; SM00082; LRRCT; 1.
DR SMART; SM00013; LRRNT; 1.
DR SMART; SM00369; LRR_Typ; 5.
KW Repeat; Signal.
FT SIGNAL 1 24 OR 23 (IN SOME ISOFORM(S)).
FT CHAIN 25 361 CHONDRODHERIN.
FT CHAIN 25 352 CHONDRODHERIN, MINOR FORM.
FT DOMAIN 79 102 10 x 24 AA LEUCINE-RICH TANDEM REPEATS.
FT REPEAT 103 102 1.
FT REPEAT 127 150 2.
FT REPEAT 151 174 3.
FT REPEAT 175 198 4.
FT REPEAT 199 222 5.
FT REPEAT 223 246 6.
FT REPEAT 248 271 7.
FT REPEAT 272 293 8.
FT REPEAT 294 317 9.
FT DISULFID 306 348 10.
FT DISULFID 308 328
FT CONFLICT 25 25 C -> Y (IN REF. 2).
FT CONFLICT 29 29 C -> W (IN REF. 2).
FT CONFLICT 31 31 C -> H (IN REF. 2).
FT CONFLICT 40 40 C -> L (IN REF. 2).
FT CONFLICT 52 52 S -> R (IN REF. 2).
SQ SEQUENCE 361 AA; 40884 MW; DA79DC98AD3DD1F8 CRC64;

```

```

Query Match 8.0%; Score 298.5; DB 1; Length 361;
Best Local Similarity 25.4%; Pred. No. 2,4e-14;
Matches 106; Conservative 64; Mismatches 152; Indels 95; Gaps 13;

```

```

QY 8 LILAWAGATATVYVWVPCPOACQIRPYTPRSSYREATVWDCNDFLTAVPAL 67
DB 6 LILXSLGLIAS--LILPALAACPCNCHC-----HSDLOHYTCXKVLQKIPKY- 51
QY 68 PACTQTLLLOSNISIVVDSOELGYLANLTELDSQNSFSDARDCEHALPQSLHLEN 127
DB 52 -----SEKTKL-----LNLQRNNPVLTATNSFRAPNIVLSLQHC 87
QY 128 QLTRLDEHSFAGIASIQELLYLNHNLRIAPAFSGLSNLRLHLSNLRALDSWFM 187
DB 88 QIREVNAAGAFRGKQIYLYLSHNDIRVLRAGAFDDELTELYLYLDHNVTELPRLGLSP 147
QY 188 LPLULEILMIGKNVDVAIILDMNFRPLANLSIVLAGNLRISDYALEGLOSLESISFYDN 247
DB 148 LVNLFITLIQLNNKIRLRSGAFGACDRLKLYLSSENSLSLOGALDVENLAKFLIDN 207
QY 248 QLARVPRALQVPGIKFLDLNKNPLQVRGPGDFANM-LHLKELGINNEELVSDKF-- 304
DB 208 QLSSTYPSAALSKLRVVEELKLSHPNPLKSPDPAFQSGFRYLETLMIDN----TLEKEFSD 263
QY 305 -ALVNPBELTKDITNNPRLSFIHPRAFHLR-----QMETLMNNLSALHQQTVS 357
DB 264 GAFLGVTTLKHVLENN-----RLHOLPSNFPFDSLETITLN----- 301
QY 358 LPLNQLQEVGLGNPIRCDCVI---RWANATGTRVREIEPOSTILCAEPPDLQRLVRE 410
DB 302 -----NPKKTCQGLRLRWLEAKTSR-----PDAT-CASPAFRGQHIND 341

```

```

RESULT 12
CBP8 HUMAN STANDARD; PRT; 536 AA.
AC P22792;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE Carboxypeptidase N 83 kDa chain (Carboxypeptidase N regulatory
DE subunit) (Fragment).
GN CPN2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=90094386; PubMed=2378615;
RA Tan F., Weerasinghe D.K., Skidgel R.A., Tamei H., Kaul R.K.,
RA Robinson I.B., Schilling J.W., Erdos E.G.;
RT "The deduced protein sequence of the human carboxypeptidase N high
RT molecular weight subunit reveals the presence of leucine-rich tandem
RT repeats."
RL J. Biol. Chem. 265:13-19(1990).
RN [2]
RP PARTIAL SEQUENCE.
RX MEDLINE=88309120; PubMed=3408501;
RA Skidgel R.A., Bennett C.D., Schilling J.W., Tan F., Weerasinghe D.K.,
RA Erdos E.G.;
RT "Amino acid sequence of the N-terminus and selected tryptic peptides
RT of the active subunit of human plasma carboxypeptidase N: comparison
RT with other carboxypeptidases."
RL Biochem. Biophys. Res. Commun. 154:1323-1329(1988).
CC -1- FUNCTION: THE 83 KDa SUBUNIT BINDS AND STABILIZES THE CATALYTIC
CC SUBUNIT AT 37 DEGREES CELSIUS AND KEEPS IT IN CIRCULATION. UNDER
CC SOME CIRCUMSTANCES IT MAY BE AN ALLOSTERIC MODIFIER OF THE
CC CATALYTIC SUBUNIT.
CC -1- SUBUNIT: TETRAMER OF TWO CATALYTIC CHAINS AND TWO GLYCOSYLATED
CC INACTIVE CHAINS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- PTM: O-GLYCOSYLATED IN THE SER/THR-RICH REGION (POTENTIAL).
CC -1- PTM: WHETHER OR NOT ANY CYS RESIDUES PARTICIPATE IN INTRACHAIN
CC BONDS IS UNKNOWN, BUT THEY DO NOT FORM INTERCHAIN DISULFIDE BONDS
CC WITH THE 50 KDa CATALYTIC SUBUNIT.
CC -1- DISEASE: A COMPLETE ABSENCE OF THE ENZYME IS NOT CONSIDERED TO BE
CC COMPATIBLE WITH LIFE.
CC -1- SIMILARITY: CONTAINS 13 LEUCINE-RICH REPEATS (LRR).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J05158; AAA51921.1; -.
DR PIR; A34901; A34901.
DR MIM; 603104; -.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000372; LRR_Nterm.
DR InterPro; IPR003591; LRR_Typ.
DR Pfam; PF00560; LRR; 11.
DR SMART; SM00013; LRRNT; 1.
DR SMART; SM00369; LRR_Typ; 10.
KW Repeat; Leucine-rich repeat; Glycoprotein.
FT NON_TER 1 1
FT REPEAT 51 74 LRR 1.
FT REPEAT 75 98 LRR 2.
FT REPEAT 99 122 LRR 3.
FT REPEAT 123 146 LRR 4.
FT REPEAT 148 170 LRR 5.
FT REPEAT 171 194 LRR 6.
FT REPEAT 196 218 LRR 7.
FT REPEAT 220 242 LRR 8.
FT REPEAT 243 266 LRR 9.
FT REPEAT 268 290 LRR 10.
FT REPEAT 291 314 LRR 11.
FT REPEAT 316 338 LRR 12.
FT REPEAT 339 362 LRR 13.

```



```

Db 330 IPDSFELHLSLCLNLSRCLRTFEARRLGSIPCLMLDLSHNALETLBLGARALGSLR 389
QY 276 -----VGPDPANMLHLKEL-----GLNNMEELY 299
Db 390 TLLGSMALNDLPYTPYANLASLORLMLQGNRVSPGCGPDEPGSCVAFSGITSLRSLX 449
QY 300 SLDK-----FALVNLPELTKLDITNNP-----321
Db 450 LYDNEIELRLBAGAFLEHFP-LTELDLSNPGLEVATGALGSLASLEVLALQNGMLAVLY 508
QY 322 -----RLSFTHPRAFHHP-----QMETLMLNNNALSLHQGVESL-PNIQEVG 365
Db 509 DLPCFLCKRLNLAEHR-LSHLPRATQAVSLLEVLDLRNNSFSLPGSAMGLETSIRLXY 567
QY 366 LHGNPIRCDCVIRW--ANATGTRVRFLEPSTJC 397
Db 568 LQGNPLSC-CGNGLMAQLHQGVADATDOLIC 600

```

```

RESULT 15
ID CHAO_TRICA STANDARD; PRT; 782 AA.
AC P82963:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chaptlin (Photoreceptor cell-specific membrane protein) (Fragment).
GN CHP.
OS Tribolium castaneum (Red flour beetle).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OX Cucujiforma; Tenebrionidae; Tribolium.
ON NCBI_TaxID=7070;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GA-1;
RA Nle W., Stronach B., Panganiban G., Brown S., Denell R.;
RT "Molecular characterization of Tc1alpha and the 3' end of the
RT Tribolium homeotic complex.";
RL Submitted (NOV-2000) to the EMBL/Genbank/DDBJ databases.
CC -1- FUNCTION: REQUIRED FOR PHOTORECEPTOR CELL MORPHOGENESIS. MEDIATES
CC HOMOPHILIC CELLULAR ADHESION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR SURFACE OF R-CELL PLASMA
CC MEMBRANE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAOPTIN FAMILY.
CC -1- SIMILARITY: CONTAINS AT LEAST 21 LEUCINE-RICH REPEATS (LRR).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, AF322227; AK01654.1; -.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR003592; LRR_out.
DR InterPro: IPR003591; LRR_typ.
DR Pfam: PF00560; LRR; 19.
DR PRINTS: PR00019; LEURICHRPT.
DR SMART: SM00370; LRR; 6.
DR SMART: SM00369; LRR_typ; 14.
KW Glycoprotein; Membrane; Repeat; Leucine-rich repeat; Vision.
FT NON_TER 1
FT REPEAT 43 66 LRR 1.
FT REPEAT 67 90 LRR 2.
FT REPEAT 93 116 LRR 3.
FT REPEAT 117 140 LRR 4.
FT REPEAT 141 164 LRR 5.
FT REPEAT 165 188 LRR 6.
FT REPEAT 224 247 LRR 7.

```

```

FT REPEAT 249 272 LRR 8.
FT REPEAT 273 296 LRR 9.
FT REPEAT 297 320 LRR 10.
FT REPEAT 321 343 LRR 11.
FT REPEAT 344 367 LRR 12.
FT REPEAT 370 391 LRR 13.
FT REPEAT 395 418 LRR 14.
FT REPEAT 419 444 LRR 15.
FT REPEAT 446 469 LRR 16.
FT REPEAT 491 513 LRR 17.
FT REPEAT 514 537 LRR 18.
FT REPEAT 539 562 LRR 19.
FT REPEAT 615 638 LRR 20.
FT REPEAT 640 663 LRR 21.
FT CARBOHYD 196 196 LRR 21.
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 454 454 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 488 488 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 530 530 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 618 618 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 648 648 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 667 667 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 782 AA; 88561 MW; 311460B2D4527917 CRC64;

```

Query Match 7.5%; Score 279; DB 1; Length 782;
 Best Local Similarity 24.8%; Pred. No. 1.9e-12;
 Matches 145; Conservative 94; Mismatches 233; Indels 112; Gaps 23;

```

QY 47 YREATVVCNDELFL-AYPALPAGTQTLLOSNSVVDSEGLYLANLELDLSSNSF 105
Db 5 HRKQFVYIGISLTLKTLHL-----SSVQNPESDAKILNREELDLSSNRL 54
QY 106 SDARDCDFHALPOLSLHLEENQTLREDSFAG--LASLQELYLNHNOLYRIAPRAFSG 163
Db 55 RNVPSNFHFLRSLKLVHLQNTITEMIHRTFGQDIIHDLTEVEYFSPNVANVQHTFAD 114
QY 164 LSNLRHLNLSNLRAIDSREFEMLPNLETIMIGKNVDALDMNFRPLANBSLYAGM 223
Db 115 LIQLEQIHLDNRRJESLERRAFMNKLSIKRLNLGNATATAYTFQNLPELEDLATYN 174
QY 224 NLREISDYALBGLQLESLSFYDN-----QLARVPRRALEQVPGI-----KFLDLNK 270
Db 175 SSSLDENIFPQVGSIG--MEHVMSHNKLNLNVAVSVPEQDTGLGQGNIKVLDLSF 232
QY 271 NPLQRVGSGDPAAM-LHLKELGLNNMEELYSIDKFAVLNPELTKLDITNNP--RLSFIH 327
Db 233 NNITSVAKQFFRPVELSLMQLYIGH-NKLNATKDLFGNMHPHLOVLDLSHNSLYELDF-- 289
QY 328 PRAFHNLPEOMETLMLNNNALSLHQGVESLPNIQEVGLHGNPIRC--DCVIRMANATGT 385
Db 290 -DTRNRKIKLQMLTSHKRISLEINDLFRPLGNLRIVDFSHNRKLSLPDMLFR---ETG- 344
QY 386 RVRFLEPOSTLCARPDLPDLVREVPFRMTDCLPLISPRSFPSLQVASESMWLHC 445
Db 345 -----LERLDVSHNLGK-----LPL-----TSLSLASQQL----- 371
QY 446 RALAPPEPEIYVTPAGLRILTPAAGR--RYRYVPESTLELRKYTABEAGLYCV--AQ 500
Db 372 -----SELDLSWNSISSL---SHGQLARFKCLSWLDLSYNRKLGQIDAQTFGIDPRLAS 422
QY 501 NLVGADTKTVSVVYGRALLQPRDEGGQLRQVGFHPHYIILSWV---PPPTVSTNLTL 557
Db 423 LNLGHNSQLLEINGLSF-----QGLTYTLHLNLDVNSLSQVPAISTPLSLSLSLA 474
QY 558 WSS-----ASSLR--GGATAPLARLPGSTHSYNTIRLL 588
Db 475 FNSLPTVALEVAGNISSLRYLNLNDLSAVPIVTHSTLSTLRHL 518

```

Sat Aug 31 14:55:23 2002

us-09-905-088a-245.rsp

Page 16

Job time: 263 sec

• • • • •

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 31, 2002, 14:36:13 ; Search time 34.24 Seconds

(without alignments)
3602.377 Million cell updates/sec

Title: US-09-905-088a-245

Sequence: 1 MRLVAPFLLAWAGATATV.....RKLPSSSEGETLLPILSONS 713

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_proteint:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriopl:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1695	45.4	730	4 Q9P231	Q9P231 homo sapien
2	1687	45.2	716	11 Q61809	Q61809 mus musculu
3	1656.5	44.4	705	4 Q43377	Q43377 homo sapien
4	1656.5	44.4	708	4 Q9H3W5	Q9H3W5 homo sapien
5	1649.5	44.2	707	11 Q9ESY6	Q9ESY6 rattus norv
6	1648.5	44.2	707	11 P97860	P97860 mus musculu
7	1625.5	43.6	718	13 Q73675	Q73675 xenopus lae
8	875.5	23.5	431	4 Q9H8Y1	Q9H8Y1 homo sapien
9	755.5	20.2	273	4 Q9NUY4	Q9NUY4 homo sapien
10	494.5	13.3	614	11 Q9D1T0	Q9D1T0 mus musculu
11	492.5	13.2	614	4 Q96F55	Q96F55 homo sapien
12	490	13.1	614	6 Q9N008	Q9N008 macaca fasc
13	468.5	12.6	606	4 Q9B220	Q9B220 homo sapien
14	442	11.8	1091	11 P70193	P70193 mus musculu
15	438.5	11.7	640	4 Q9HCJ2	Q9HCJ2 homo sapien
16	432.5	11.6	540	5 Q9V053	Q9V053 drosophila

17	431	11.5	1094	4 Q9BVB8	Q9BVB8 homo sapien
18	430	11.5	1093	4 Q9BVB1	Q9BVB1 homo sapien
19	427.5	11.5	532	5 Q96671	Q96671 drosophila
20	420.5	11.3	628	4 Q9BT80	Q9BT80 homo sapien
21	407	10.9	733	5 Q24250	Q24250 drosophila
22	407	10.9	737	5 Q9VU51	Q9VU51 drosophila
23	406.5	10.9	719	4 Q96NT6	Q96NT6 homo sapien
24	401.5	10.8	737	5 Q965M3	Q965M3 caenorhabdi
25	401.5	10.8	881	5 Q965M2	Q965M2 caenorhabdi
26	401.5	10.8	1447	5 Q16779	Q16779 caenorhabdi
27	399.5	10.7	811	4 Q75139	Q75139 homo sapien
28	399	10.7	78	11 Q61974	Q61974 mus musculu
29	398.5	10.7	789	6 Q9BE71	Q9BE71 macaca fasc
30	395.5	10.6	673	11 Q9C2T5	Q9C2T5 mus musculu
31	392.5	10.5	792	4 Q9UDT7	Q9UDT7 homo sapien
32	388.5	10.4	832	4 Q9ULH4	Q9ULH4 homo sapien
33	386	10.3	492	11 Q99KT6	Q99KT6 mus musculu
34	386	10.3	653	4 Q9H8Y1	Q9H8Y1 homo sapien
35	385	10.3	788	4 Q9CYK3	Q9CYK3 mus musculu
36	378	10.1	649	4 Q96A85	Q96A85 homo sapien
37	377	10.1	1065	4 Q94898	Q94898 homo sapien
38	374.5	10.0	700	4 Q9P244	Q9P244 homo sapien
39	368	9.9	907	11 Q921P4	Q921P4 mus musculu
40	361.5	9.7	542	5 Q9N4G5	Q9N4G5 caenorhabdi
41	361.5	9.7	809	11 Q9DBY4	Q9DBY4 mus musculu
42	361	9.7	1173	5 Q9V7J8	Q9V7J8 drosophila
43	360.5	9.7	1531	11 Q88279	Q88279 rattus norv
44	360	9.6	473	11 Q99P18	Q99P18 mus musculu
45	360	9.6	907	4 Q75473	Q75473 homo sapien

ALIGNMENTS

RESULT	ID	Q9P231	PRELIMINARY:	PRT:	730 AA.
Q9P231	Q9P231	Q9P231	Q9P231		
AC	Q9P231	Q9P231	Q9P231		
DT	01-OCT-2000	(TREMBLrel. 15, Created)			
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)			
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)			
DE	KIAA1497	PROTEIN (FRAGMENT).			
GN	KIAA1497				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=20277482; PubMed=10819331;				
RA	Nagase T., Kikuno R., Ishikawa K., Hirosewa M., Ohara O.;				
RT	genes.XVII. The complete sequences of unidentified human				
RT	RT which code for large proteins in vitro.;				
RL	DNA Res. 7:143-150(2000).				
DR	EMBL: AB040930; BAA96021.1; -				
DR	InterPro: IPR003961; FN.III.				
DR	InterPro: IPR003598; IG.C2.				
DR	InterPro: IPR003006; IG.MHC.				
DR	InterPro: IPR001611; LRR.				
DR	InterPro: IPR000483; LRR_Cterm.				
DR	InterPro: IPR00372; LRR_Nterm.				
DR	InterPro: IPR003592; LRR_out.				
DR	InterPro: IPR003591; LRR_Typ.				
DR	Pfam: PF00047; Ig. 1.				
DR	Pfam: PF00560; Ig. 9.				
DR	Pfam: PF01463; LRRCT. 1.				
DR	Pfam: PF01462; LRRNT. 1.				
DR	SMART: SM00060; ENS. 1.				
DR	SMART: SM00408; IGC2. 1.				
DR	SMART: SM00370; LRR. 4.				
DR	SMART: SM00082; LRRCT. 1.				
DR	SMART: SM00013; LRRNT. 1.				


```

Oy      619 --TSCFRALGDRPGLIAMIILAAVLL 641
       |: :| :||::||:
Db      626 HETSTALAAMGSMFAVISLAIATYI 652

RESULT      3
043377      PRELIMINARY;          PRT;           705 AA.
AC      043377;
DT      01-JUN-1998 (TREMBLrel, 06, Created)
DT      01-JUN-1998 (TREMBLrel, 06, Last sequence update)
DT      01-DEC-2001 (TREMBLrel, 19, Last annotation update)
DE      PROBABLE LEUCINE-RICH REPEAT PROTEIN.
GN      RGI1BD07.1.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBL_Taxid=9606;
RN      [1]
RE      SEQUENCE FROM N.A.
RC      TISSUE=FIBROBLAST;
RA      Murray J., Langston Y., Clarke C.;
RL      Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
RR      EMBL: AC004142; AAC02752.1; -
DR      InterPro: IPR003961; FN_III.
DR      InterPro: IPR003588; Ig_C2.
DR      InterPro: IPR003006; Ig_MHC.
DR      InterPro: IPR001611; LRR.
DR      InterPro: IPR000483; LRR_Cterm.
DR      InterPro: IPR000372; LRR_Nterm.
DR      InterPro: IPR003592; LRR_out.
DR      InterPro: IPR003591; LRR_typ.
DR      Pfam: PF00041; fn3; 1.
DR      Pfam: PF00047; Ig_1.
DR      Pfam: PF00560; LRR; 9.
DR      Pfam: PF01463; LRCT; 1.
DR      Pfam: PF01462; LRNT; 1.
DR      PRINTS: PR00019, LEURICHRPT.
DR      SMART: SM00408; IGc2; 1.
DR      SMART: SM00370; LRR; 4.
DR      SMART: SM00082; LRCT; 1.
DR      SMART: SM00013; LRNT; 1.
DR      SMART: SM00369; LRR_typ; 1.
KW      Immunoglobulin domain.
SQ
SEQUENCE   705 AA; 79049 MW; 9846066693FABCIA CRC64;

Query Match              44.4%; Score 1656.5; DB 4; Length 705;
Best Local Similarity    47.2%; Pred. No. 6e-118;
Matches 333; Conservative 109; Mismatches 217; Indels 47; Gaps
5.

Oy      17 TATPVVPMHVPDPCACQIRWYTTPRSSYREATTYDCNDLFTAVPALPGTOTILL 76
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      14 TTLVAQAVDKKVCDCRCLCTCEIRMPETPSIYMESATYDCNDLGITPPALPANTQILL 73
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Oy      77 QNSIAYRVQDSGLGIANTLTEDLSQNSFSAPDCDFHALPOLISLHLEFNQLTREDHS 136
       |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      74 QTNNIKAIKYS-TDPENVLTGDLSDQNMLSSVTNINWKMKPOLSYLVLENKLTPEPKC 132
       |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Oy      137 FAGLASLOEYLNHNOLVRIPARPSGSNLRLHMSNLRLRADSRWFEMLPNELIIMI 196
       |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      133 LSELISMLOELYTNHNLSTISGAFTGLHNLRLHLSNRLOMNSKMFADPNLEIIMI 192
       |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Oy      197 GGKKVAAILDMNRPRIANLRSLVLAAGMLREISYALEGLQSLESSEFYDNQCARVPRA 256
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      193 GENPIIRIDMNFKLINRSRYVINGINTLTIPDNALVGLENSESIFYNRIKXPHYA 252
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Oy      257 LEQVPEGLAKFLDNKNPDLOVGPDEFANNLHLKEGLGNMEELVSIDKFALYNLPETLKID 316
       ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      253 LQKVVMLEKFLDNKNPINIRIGCPFSNMLHLKEGLGINNMPELISIDSLADVNPDLRKIE 312
       ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Oy      317 ITNNPLSLTHRARHHLPOMETLMINNNAALAHQOTVESLRLQDEVGLHGVPITCDGV 376
       |||||||::|::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

Db	313	ANNNRRLSYIHNNAFRLPKLESLMNSNALMSALYHGHTESLPMLKELSHSNPIRCDV	372
Qy	377	IRMANATGRVRFETIPQSTLCAERPDQLRPLVREVPREMTDCLPLISPRSPSLQVA	436
Db	373	IRMMNNKNTNIRFEMPPDLFCVDPPEFGQNVROVHFHDMWEICLPLIAESPSPNLNE	432
Qy	437	SESENVLCRRLAEPEEPIYWTATAGLRLTPAHAGRRKRVYPECTELRVTAEALYT	486
Db	433	AGSYVSEFCRATAEQPEPIYWTITPSGQGLLPTNLTDKRFYHSEETLDINGTPEKGLYT	492
Qy	497	CYAMNLVADRTKTVSVVVGRRLLDPGRDEGQGLELRQETHPIHLLSWYPRPTVSTNL	556
Db	493	CIATLPLVGADKLSVKMIKVDGSEFPQ---DNNSLNKIRIDIOANSVLYSMKASSITLSSV	549
Qy	557	TWSASLSLGGCAVALARLPRGTHSYNTLRLLQATEYMACLQVAFADAHQTOLACWARTK	616
Db	550	KWTAFVKTENSHAQASARIPSDVKVYNTLHLPSTEXKICIDIPYIKNRKKCVANTYTK	609
Qy	617	-----EATSCRRALGDRPGLIALLALLVLLAAGLAAHLTGCPKRGVGGRR- 663	657
Db	610	GLHPDQKEYEKNNNTTLMACGLGLIGIVCLLSCL-----SPENNCDSGHS 657	657
Qy	664	-----PLPPANAFMWSAPVSPVRSAPLVRPWN 691	691
Db	668	YVRNVLQKPTFALGELYPPLINLMDEAGKEKSTSLKVAIYIGLPTN 703	703
RESULT	4		
Q9H3W5			
AC	09H3W5	PRELIMINARY;	PRT; 708 AA.
DI	01-MAR-2001 (TREMBLrel.16, Created)		
DT	01-MAR-2001 (TREMBLrel.16, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel.19, Last annotation update)		
DE	HYPOTHETICAL 79.4 KDA PROTEIN (NEURONAL LEUCINE-RICH REPEAT		
DE	PROTEIN-3).		
GN	DRKP2761K424 OR NLR-3.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
OX	NCBI_Taxid=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=AMIGDALA;		
RA	Bloecker H., Boecker M., Brandt P., Mewes H.W., Well B., Wiemann S.;		
RL	"Human neuronal leucine-rich repeat protein-3(NLR-3)".		
RL	Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AL442092; CAC09450.1; -		
DR	EMBL; AB060967; BAB47184.1; -		
DR	InterPro; IPR003961; FN_III.		
DR	InterPro; IPR003599; Iq.		
DR	InterPro; IPR003598; Iq_c2.		
DR	InterPro; IPR003006; Iq_MHC.		
DR	InterPro; IPR001611; LRR.		
DR	InterPro; IPR000483; LRR_Cterm.		
DR	InterPro; IPR000372; LRR_Outem.		
DR	InterPro; IPR003592; LRR_cult.		
DR	InterPro; IPR003591; LRR_tyr.		
DR	Pfam; PF00041; fn3.1.		
DR	Pfam; PF00047; Iq.1.		
DR	Pfam; PF00560; LRR.9.		
DR	Pfam; PF01463; LRRCT.1.		
DR	Pfam; PF01462; LRRNT.1.		
DR	PRINTS; PR00019; LEURICHRPT.		
DR	SMART; SM00060; FN3.1.		
DR	SMART; SM00409; IG.1.		
DR	SMART; SM00408; IGC2.1.		
DR	SMART; SM00370; LRR.4.		
DR	SMART; SM00082; LRRCT.1.		

Db 483 IKGITPKEGGTLCIATNLVYGADLKSIMIKVGGFVPO---DNNGSLNLIKIDIRANSVLY 539
 Qy 544 SVWTPPNTSTYNTLWSSASLKGQATALARLPRGTHSYNTRILLQATEYMACLOAVAD 603
 Db 540 SKKASKILKSSVKVTAIVKTEDSOAQSARIPSDVKYKYNLTHLKPSTEVKICIDIPITY 599
 Qy 604 AHYOLACYWART-----KEATSCS---RALGDRPLLIALALAVLLAAGLAHLGT 652
 Db 600 QSRKQCVAVVTTKSLHDGKENGKSHTEFVACVGLLIGVMCL-----FGCVSQSGN 653
 Qy 653 GQPRKGVGRRLPPAMAFWCHSAPSVYASAPLYLPNNPGRKLPKRSSE 701
 Db 654 CENEHSYTVNHCKPTLAF-----SELYPPLINIMESSEKRPASLE 694

RESULT 6
 P97860 PRELIMINARY; PRT; 707 AA.
 AC P97860;
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE LEUCINE-RICH REPEAT PROTEIN PRECURSOR (FRAGMENT).
 GN LRRR3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96233817; PubMed=9011764;
 RA Taniuchi H., Tohyama M., Takegi T.;
 RT "Cloning and expression of a novel gene for a protein with leucine-
 RT rich repeats in the developing mouse nervous system.";
 RL Brain Res. Mol. Brain Res. 36:45-52(1996).
 DR EMBL; D49802; BAA08622.1; -.
 DR MGI; MGI:106036; Lrrn3.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR003598; Ig_c2.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR000483; LRR_Cterm.
 DR InterPro: IPR000372; LRR_Mterm.
 DR InterPro: IPR003592; LRR_out.
 DR InterPro: IPR003591; LRR_tyr.
 DR Pfam; PF00041; fn3; 1.
 DR Pfam; PF00047; Ig; 1.
 DR Pfam; PF00560; LRR; 10.
 DR Pfam; PF01463; LRRCT; 1.
 DR Pfam; PF01462; LRRNT; 1.
 DR PRINTS; PR00019; LEURICHRPT.
 DR SMART; SM00060; FN3; 1.
 DR SMART; SM00408; IGC2; 1.
 DR SMART; SM00370; LRR; 4.
 DR SMART; SM00082; LRRCT; 1.
 DR SMART; SM00013; LRRNT; 1.
 DR SMART; SM00369; LRR_tyr; 1.
 KW Immunoglobulin domain; Signal.
 FT SIGNAL 1 20
 FT CHAIN 21 >707 POTENTIAL.
 FT NON_TER 707
 SQ SEQUENCE 707 AA; 79156 MW; FAETCL573DDDL65B CRC64;

Query Match 44.2%; Score 1648.5; DB 11; Length 707;
 Best Local Similarity 47.7%; Pred. No. 2.5e-117;
 Matches 328; Conservative 111; Mismatches 225; Indels 23; Gaps 5;

Oy 17 TATVVVVMWHPCCPOGCAQIRPWTPSSYREATVVCNDLFTLAVPALPAGNQTLL 76
 Db 17 TTVLVAVIKKKVDCPOLCTCEIRPMTTPRSIYMEASTVDCNDLGLNFPARLPADTQILL 76

Qy 77 QSNISIVRDSQSELGYLANLTLDLSQNSFSDARDCDFHALPOLLSTLHEENQTLREDS 136
 Db 77 QTNNIARIIES-TDFPVNLGTDLDSQNNLSSVINIVQKSSQLSYLEENKLTLEPERC 135
 Qy 137 FAGIASIQEELYNHNOYLRIAPAFSGLSNLLHLHNSNLLRAIDSRWEMLENIIM 196
 Db 136 LYGSLNIOELLYVNNHLLSTISPAFIOIHLNLRLHNSNLLQMINSGWDPALNLEIM 195
 Qy 197 GGNKVDAIILMNPFRANLRSIYLAGNNLREISDYALEGQSIESPSFYDNOIARPRRA 256
 Db 196 GDNPIIRIKDMNQPLKLSLVYAGINTLEIPDDALAGLENEISIFDNRLSKVPQYA 255
 Qy 257 LEQVPLKFLDLNKNPLQRYGPDGFANMLKELGLNNMEVLSIDKFAVNLPELTID 316
 Db 256 LQKAVNLKFLDLNKNP-INRIIRGDFSNMLHLELGINNMEVLSIDSLAVDNLPLRKIE 315
 Qy 317 ITNPRLSFTIHPRAFHLPMOETLMNNALSLAQOYVSLDNLQVGLRGPIPCDY 376
 Db 316 ATNPNRLSYIHPNAFFRLPRLTESLMTNLSALYHGTIESLPNLKEISHSNPIRCDCV 375
 Qy 377 IRMANATGTRVRFIEPOSTLCAEPPDLQRLPVEVEPPREMTDCHLPLISPRSPSLQYA 436
 Db 376 IRWIMNMTKINIREPDSLCVDPPEFGQONVQVHFQDMETCLPLIAPESPDLDB 435
 Qy 437 SGEWVLHCRALAEPEPELIYVTPAGILTPAHAGRRYRYPBGTDELRRVTAEGALY 496
 Db 436 ADSYSLHCRATAPQGEIYVITPSGKLLPNTMREKIFYHSGTLEIRGTPEGGL 495
 Qy 497 CVANLVGADTKYVSVYVGRALLQPGDEGQELRQVQEHPIHLLSWTPNTYSTNL 556
 Db 496 CIATNLVGAULKSIMIKVGGVPO---DNNGSLNLIKIDIRANSVLYSKASKILKSV 552
 Qy 557 TWSSASSLRQCATALARLPRGTHSYNTRILLQATEYMACLOVAFDAHTOLACYWART 616
 Db 553 KWICFVTEDESHAQSAKRIPSDVKYKYNLTHLKPSTEVKICIDIPITYQSRKCCVWTK 612
 Qy 617 EATSCHRALDR-----PGLIALIALAVLLAAGLAHLGTGQPRKGVGRRLPPA 668
 Db 613 SLEHDGKEYGNHNVFACVGLLIGVMCLF---SCVSQESSSGEHSYANNHCXRA 669
 Qy 669 WAEWMSAPSVRYVAPLYLPNNPGRK 695
 Db 670 LAF-----SELYPPLINIMESSEK 688

RESULT 7
 ID 073675 PRELIMINARY; PRT; 718 AA.
 AC 073675;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE NEURONAL LEUCINE-RICH REPEAT PROTEIN.
 GN XNLR-1.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99070063; PubMed=9852961;
 RA Hayata T., Uochi T., Asashima M.;
 RT "Molecular cloning of XNLR-1, a Xenopus homolog of mouse neuronal
 RT leucine-rich repeat protein expressed in the developing Xenopus
 RT nervous system.";
 RL Gene 221:159-166(1998).
 DR EMBL; AB014462; BAA28530.1; -.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR003598; Ig_c2.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR000483; LRR_Cterm.

```
DR InterPro: IPR000372; LRR_Nterm.
DR InterPro: IPR003592; LRR_out.
DR InterPro: IPR003591; LRR_Typ.
DR Pfam: PF00041; fn3; 1.
DR Pfam: PF00047; fn3; 1.
DR Pfam: PF00560; LRR; 9.
DR Pfam: PF01463; LRRCT; 1.
DR Pfam: PF01462; LRRNT; 1.
DR PRINTS: PR00019; LEURICHRPT.
DR SMART: SM00060; FN3; 1.
DR SMART: SM00408; IGC2; 1.
DR SMART: SM00370; LRR; 3.
DR SMART: SM00082; LRRCT; 1.
DR SMART: SM00013; LRRNT; 1.
DR SMART: SM00369; LRR_Typ; 3.
KW Immunoglobulin domain.
SQ
SEQUENCE 718 AA; 80035 MW; AEP840DCB28B837 CRC64;

Query Match 43.6%; Score 1625.5; DB 13; Length 718;
Best Local Similarity 52.1%; Pred. No. 1.4e-115;
Matches 335; Conservative 98; Mismatches 201; Indels 9; Gaps 6;

OY 3 LLYAPLLANVAGATATVPVPMHVPQPCACQIRPWTTPNSSTREATTVDGNDLFLTA 62
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
DB 8 VLTACQLLGLITSLTSDSSAPAN-ECPOLCVCERMPFTPOSTYREATTVDGNDLRTK 66
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
OY 63 VPPALPAGTQTLTLLQNSIYRVDSQELGYANTTELDSNFSRSDARDCDFHAPOLSL 122
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
DB 67 IPANLSIDTQVLLQSNINIKTN-GEIQRLVNTTELDSQNNNTSHDVGSLSSQITTL 125
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
OY 123 HLEENQITLEDHSFAGLSLQELYLNHNOYKRIAPRAFSGLSNLLRLHNSNLTARIDS 182
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
DB 126 HLEENQIFETWTDYCLDITLQELYNHNOINSISPAFSGLNLLRLHNSNLTARIDS 185
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
OY 183 RWEEMLPNLEITLIGNKVDALIDMNPRLANRSLVLAQMLREISDVALBSQLESL 242
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
DB 186 RWEESTPNLEITLIGENPVYGLIDLNFOPLVNLRSVLGMYTSDIPGNALGLDML 245
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
OY 243 SFYDNOIARVRRALRQVPLKFLDLINKNPLOKRVGPDGFANMLHELGINNNEELV 302
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
DB 246 SLVDNKGKVPOLALQKVPRLKFLDLINKNPRIQSGDFKNMKTKELGGINNAELV 305
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
OY 303 KRLVNLPELTLDITNNPRLSTIPRAFHILPOMETLMNNALSALHQVYESLPNLQ 362
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
DB 306 RNAMENLPETLKEATNPNKLSYIHRSAFRNVPTELSIMNNALNSVVRGVESE 365
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
OY 363 EVGLHGPIKDCQVIRMANATGRVRFIEPOSTLCAEPDLOLRPVREVPFRMT-DH 421
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
DB 366 EISTHSPRLKDCVLIHMGSNOTSIRMEPLSMFCALPEYRQGPVKALADDPAGE 425
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
OY 422 PLISPRSPPSLQVAGSESVLHCRALAEPEPIYVWTPAGLRTPAHAGRRYRVPE 481
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
DB 426 PMISQDFPHSLDIDGMTSLDCCRADAPEPEIYVWTPGLGHVTLLETJSDKHL 485
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
OY 482 LEIARVTAEBAGLYTCVAQNLVGADTKTVSVVVGALLQPGRDEGGLELRVQ 541
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
DB 486 LQIFNVQVEDSGRYTCVAQNSEGADTKVATLRVNGTLL----DGTQALRL 541
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
OY 542 LLSWVPNPNVTNLTWSSAS-SLRQOGATALARLRGTHSVYITLLDQATEYMAC 600
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
DB 542 LVSWMKVSSTVSLNLMKSSATMKIDNPHTIYARVADVEHYNLTLQPALEYE 601
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
OY 601 FADAHTQACVMARKEATSCHRALGDRPGLIILALAVLLA 643
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
DB 602 GLHQQAQRACINVTTK-GTISYSLTVTQDETSAALAAVMSLFA 643
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |

RESULT 8
O9H8V1 PRELIMINARY; PRT; 431 AA.
AC O9H8V1;
DT 01-MAR-2001 (Tremblrel, 16, Created)
```

```
DT 01-MAR-2001 (Tremblrel, 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel, 19, Last annotation update)
DE CDNA FLJ13211 FIS, CLONE NT2RP4000907, HIGHLY SIMILAR TO MOUSE NLR-1
DE LEUCINE-RICH-REPEAT PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;
RA "NEDO human cDNA sequencing project.";
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AK023273; BAB14500.1; -.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003599; Ig.
DR InterPro: IPR003598; Ig_C2.
DR InterPro: IPR003600; Ig_Like.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR000483; LRR_Cterm.
DR Pfam: PF00047; Ig; 1.
DR Pfam: PF00560; LRR; 1.
DR Pfam: PF01463; LRRCT; 1.
DR SMART: SM00060; FN3; 1.
DR SMART: SM00409; IGC; 1.
DR SMART: SM00408; IGC2; 1.
DR SMART: SM00410; IGLike; 1.
DR SMART: SM00082; LRRCT; 1.
DR SMART: SM00369; LRR_Typ; 1.
KW Immunoglobulin domain.
SQ
SEQUENCE 431 AA; 48623 MW; 999161E63B894B CRC64;
```

```
Query Match 23.5%; Score 875.5; DB 4; Length 431;
Best Local Similarity 46.9%; Pred. No. 1.3e-36;
Matches 173; Conservative 69; Mismatches 110; Indels 17; Gaps 3;

OY 284 MLHKEELGNMNEELVSDIKFALVNLPELTLDITNNPRLSTIPRAFHILPOMETLM 343
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
DB 1 MRLKEELGNMNEELVSDIKFALVNLPELTLDITNNPRLSTIPRAFHILPOMETLM 60
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
OY 344 NNALSALHQVYESLPNLQELYLNHNOYKRIAPRAFSGLSNLLRLHNSNLTARIDS 403
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
DB 61 NNALNAIYQKTVESLPNLRREISHSNPLRCDCVIHMINSKNTNIRFMEPLSMFC 120
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
OY 404 QRLPVREVPFRMDHCLPLISPRSPPSLQVAGSESVLHCRALAEPEPIYVWTPAG 463
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
DB 121 KGHQKEVLIDQSSBQCLPMISHDSFPMRLNVDTGTTFVLDCCRAMAEPEPIYV 180
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
OY 464 RLTPAHAGRRYRVPEGLRLRYTAEBAGLYTCVAQNLVGADTKTVSVVVGALLQ 523
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
DB 161 KITVETLSDKTKLISSEGLTLEISNIOIEDSGRTYTCVAQNVGADTRVATIKY 236
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
OY 524 DEGGLELRVQETDHPHYILSWVPNPNVTNLTWSSAS-SLRQOGATALARLRGTH 582
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
DB 237 DGTQVLYKTYVQTESHSLSVMKVSNTNLMKSSATMKIDNPHTIYARVADVEH 296
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
OY 583 NITRLQATEYMACLOVAFAHAHTQACVMARKEA-----TSCHRALGDRPG 630
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
DB 297 NLTHTLPSTDEYVCLTVSNIHQOTOKSCVNVNTRNAFAVDISDETSTALAAV 356
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
OY 631 LIATLALAV 639
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
DB 357 VISLSTAV 365
   :| | | | | : : : | | | | | : : : | | | | | : : : | | | | |

RESULT 9
O9NU04 PRELIMINARY; PRT; 273 AA.
AC O9NU04;
DT 09NU04;
```


Query Match	Similarity	Score	DB	Length
Best Local Similarity	58.8%	Pred. No. 1e-49;		
Matches 151;	Conservative	39;	Mismatches 66;	Indels 1;
				Gaps 1

ID	OSD1T0	PRELIMINARY;	PRT;	614 AA.
AC	09D1T0;			
DT	01-JUN-2001 (Tremblrel. 17, Created)			
DT	01-JUN-2001 (Tremblrel. 17, Last sequence update)			
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)			
DE	ADULT MALE TESTIS CDNA, RIKEN FULL-LENGTH ENRICHEDB LIBRARY,			
DE	CLONE:4930471K13, FULL INSERT SEQUENCE.			
OS	Mus musculus (mouse).			
OC	Eumetazoa: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:			
OC	Mammalia: Eutheria: Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	(1)			
RP	SEQUENCE FROM N.A.			

ID	OSD1T0	PRELIMINARY;	PRT;	614 AA.
AC	09D1T0;			
DT	01-JUN-2001 (Tremblrel. 17, Created)			
DT	01-JUN-2001 (Tremblrel. 17, Last sequence update)			
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)			
DE	ADULT MALE TESTIS CDNA, RIKEN FULL-LENGTH ENRICHEDB LIBRARY,			
DE	CLONE:4930471K13, FULL INSERT SEQUENCE.			
OS	Mus musculus (mouse).			
OC	Eumetazoa: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:			
OC	Mammalia: Eutheria: Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	(1)			
RP	SEQUENCE FROM N.A.			

```

Query Marc30 13.38: Score 494.5; DB 11; Length 614;
Best Local Similarity 25.68: Pred. No. 2.9e-29;
Matches 173; Conservative 97; Mismatches 251; Indels 155; Gaps 18.

QY      1 MRLIVADLLIAM-----VAGATATVPVPMHVPQPCACQAIRPMYTPRSSYREA 50
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      6 MRSMSPRLAACWQRIILLVIGSLSSAT-----GCPRCES-----AQC 46

QY      51 TTVDCNDLFLAVPPALPACTOTILLQNSIYRVQDSGLYANITELDLSQNSFSFARD 110
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      47 RAVLCHRRREYAVEGIPETTRILLDGKRRIRITLQDQEFASPHLEDELENNIVSAVED 106

QY      111 CDFH-----ALPQLSLHLEENQDRLREHDSFAGISAOEL 146
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      107 GAFNNLEENLRTLGIRSNRKLPLGVFTGLSNLTKLIDISENKIYILLDYMFODIYINAKSL 166

QY      147 YLNNNOYLRIAPRAFSGLSMLRLAHNSMLRAIKDSRPFEMPLNLEITMIGNKVDAIILD 206
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      167 EYGDNDLVYISHRAFSGLSNLEBQLETKKNLISFTEALSHDHGIVRLRLHNLINAIRD 226

QY      207 MNEPRLANLSVLAGN-NIREISDALEGLQSLSESFYNDQARVPARRALEOVPELKT 265
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      227 YSFRLRYRLKVLIEISHMYPIDYTPMCLYGL-NLTSLSITHCNLTAVPYAVLRHLYLRF 285

QY      266 LDLMKNPQVAVGRODFANMLHLKELGSLNNKMELVASIDKFAVYNLPBELTKDITNNRPLSF 325
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      286 LNLSTNPGITGTE---GSMH-----ELRLRQETQV-----GGGLAV 319

```

```

QY 326 IHPRAFHHLPEMETLMLNNALSAHQVTESLPNLOEVLHGPNIRDCVIRMANATGT 385
D 320 VEPYAFKRLNLRVNLVSGNQLTLEESVFSVGNLETLILDSNPLACDCRLMWFRRRW 379
QY 386 RVRFIEPQSTLCAEPDQLRLVREVPFREMTHCLPLISPRSF-----PPLSQ 434
D 380 RLNFNRQOPT-CATPEFVQGEKEDFP-----DVLFPNFTCCRARIIDRKAQOVF 429
QY 435 VASGESMVLHCRALAEPEPEIYWTTPAGLRITPAHAGRRYVPEGTLELRVTAEEAGL 494
D 430 VDEGHTVQFCRADGDPPIALMLSPR-KHLVSAKSNGLTVFPGTLEVRAYQVODNGT 488
QY 495 YTCVAQNLVGAADTKTVSVVVGRRALQPGRDEGQGLELRVQETHPHYHLLSWTPPNVTST 554
D 489 YLCIAANAGGNDM-----PAHLHVSYSPPMHP 519
QY 555 NLTWSSASSLRGGATALARPRT--HSYNITRLQATE-----YMACLOVAFAAD 603
D 520 NKTFAFISNOPEGE--ANSTRATVPPFDIKTLIATMGFISFLGVVLCVILF-- 574
QY 604 AHTQLACWARKTEAT 619
D 575 -----LMSRGKNT 583

```

```

RESULT 11
Q96FE5 PRELIMINARY: PRT: 614 AA.
AC Q96FE5:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE UNKNOWN (PROTEIN FOR MGC:17422).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=BRIN, N.A.
RC TISSUE=BRIN, N.A.
RA Strusberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC011057; AAH1057.1; -.
SEQUENCE 614 AA: 69145 MW; EFD967E3B716698D CRC64;

```

```

Query Match 13.2%; Score 492.5; DB 4; Length 614;
Best Local Similarity 25.4%; Pred. No. 4,1e-29;
Matches 172; Conservative 98; Mismatches 251; Indels 155; Gaps 18;

```

```

QY 1 MRLVAPILLAW-----VAGATATVPVPMHVPCPCACQIRPWTPRSSYREA 50
D 6 VRMPSPILACWOPILLVIGVLSGSAI-----GCPGCCS-----AQD 46
QY 51 TTVDCNDLFTAVPALPAGTQTLQSSIVRVDSOSELGYANTLDELSONSFSQD 110
D 47 RAVLCGRKRVAVVEGIPETRLDGLGRKRIKTLNODFASPHLELNINIVSAVEP 106
QY 111 CDH-----ALPOLSHLENDQTRLEDHSPFAGLSLOEL 146
D 107 GANNMLFNTLTGLRSNRKLIPLGVFTGISNTFKLIDSENKIVILLDYFODLNLKSL 166
QY 147 YLHNDLYRIAPAFSGLSNLRHLNSMLRAIDSRKWEMLPNLEILMIGKKVDAIID 206
D 167 EVDGNDLVYISHAFSGLSNLEBDLTKCLNLTIPREALSHLGLVLLRHNINAIAD 226
QY 207 MNRPLANRSLVLAGM-NIREISDYALEGSLSESLFYDNOLANVRRALEQVPGLEF 265
D 227 YSRKRLVRLKVLLEISHWPYLDTPNCLVGL-NLTSLSITHCLVAVPYLAVALHLYLRF 285
QY 266 LDKNRPLOVVGPGDFANMHLKELGLNNMEELVSDKFAVLVLPRLTKLIDITNNRSLF 325

```

```

D 286 LNSTNPITIE-----GSMH-----ELLRQELQV-----GGQLAV 319
QY 326 IHPRAFHHLPEMETLMLNNALSAHQVTESLPNLOEVLHGPNIRDCVIRMANATGT 385
D 320 VEPYAFKRLNLRVNLVSGNQLTLEESVFSVGNLETLILDSNPLACDCRLMWFRRRW 379
QY 386 RVRFIEPQSTLCAEPDQLRLVREVPFREMTHCLPLISPRSF-----PPLSQ 434
D 380 RLNFNRQOPT-CATPEFVQGEKEDFP-----DVLFPNFTCCRARIIDRKAQOVF 429
QY 435 VASGESMVLHCRALAEPEPEIYWTTPAGLRITPAHAGRRYVPEGTLELRVTAEEAGL 494
D 430 VDEGHTVQFCRADGDPPIALMLSPR-KHLVSAKSNGLTVFPGTLEVRAYQVODNGT 488
QY 495 YTCVAQNLVGAADTKTVSVVVGRRALQPGRDEGQGLELRVQETHPHYHLLSWTPPNVTST 554
D 489 YLCIAANAGGNDM-----PAHLHVSYSPPMHP 519
QY 555 NLTWSSASSLRGGATALARPRT--HSYNITRLQATE-----YMACLOVAFAAD 603
D 520 NKTFAFISNOPEGE--ANSTRATVPPFDIKTLIATMGFISFLGVVLCVILF-- 574
QY 604 AHTQLACWARKTEAT 619
D 575 -----LMSRGKNT 583

```

```

RESULT 12
Q9N008 PRELIMINARY: PRT: 614 AA.
AC Q9N008:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOHETICAL 69.2 KDA PROTEIN.
OS Macaca fascicularis (Crab eating macaque) (Cynomolpus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheidae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP TISSUE=CEREBELLUM CORTEX;
RC TISSUE=CEREBELLUM CORTEX;
RA Osaka N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB046639; BAB03557.1; -.
DR HSSP; P23945; IXUN.
DR InterPro: IPR003598; Iq_C2.
DR InterPro: IPR003006; Iq_MHC.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR000483; LRR_Cterm.
DR InterPro: IPR000372; LRR_Nterm.
DR InterPro: IPR003592; LRR_out.
DR InterPro: IPR003591; LRR_tyr.
DR Pfam: PF00047; Iq_1.
DR Pfam: PF00560; LRR_9.
DR Pfam: PF01463; LRRCT_1.
DR Pfam: PF01462; LRRNT_1.
DR SMART; SM00408; IqC2; 1.
DR SMART; SM00370; LRR_5.
DR SMART; SM00082; LRRCT_1.
DR SMART; SM00013; LRRNT_1.
DR SMART; SM00369; LRR_tyr; 1.
KW Hypothetical protein; Immunoglobulin domain.
SEQUENCE 614 AA: 69187 MW; BA6CB8C7C993BE9A CRC64;

```

```

Query Match 13.1%; Score 490; DB 6; Length 614;
Best Local Similarity 26.4%; Pred. No. 6,4e-29;
Matches 160; Conservative 89; Mismatches 229; Indels 128; Gaps 16;

```


AC P70193;
 DT 01-FEB-1997 (TrEMBLrel. 02, Created)
 DT 01-FEB-1997 (TrEMBLrel. 02, last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)
 DE MEMBRANE GLYCOPROTEIN.
 GN IMG.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10990;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96394313; PubMed=8798419;
 RA Suzuki Y., Sato N., Toyama M., Wanaka A., Takagi T.;
 RT "cDNA cloning of a novel membrane glycoprotein that is expressed
 specifically in glial cells in the mouse brain LIG-1: A protein with
 leucine-rich repeats and immunoglobulin-like domains.";
 RL J. Biol. Chem. 271:22522-22527(1996).
 DR EMBL: D78572; BAA11416.1; -
 DR HSSP: P56276; ITLK.
 DR MGD: MGI:107935; Img.
 DR InterPro: IPR003598; Ig_C2.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR000483; LRR_Cterm.
 DR InterPro: IPR000372; LRR_Nterm.
 DR InterPro: IPR003592; LRR_out.
 DR Pfam: PF00047; Ig_3.
 DR Pfam: PF00560; LRR_14.
 DR Pfam: PF01463; LRRC1.
 DR Pfam: PF01462; LRRC1.
 DR SMART: SM00408; IgC2; 3.
 DR SMART: SM00370; LRR; 6.
 DR SMART: SM00082; LRRC1; 1.
 DR SMART: SM00013; LRRC1; 1.
 DR SMART: SM00369; LRR_Typ; 4.
 KW Immunoglobulin domain.
 SQ SEQUENCE 1091 AA; 119283 MW; A13D0866CE4C203D CRC64;

Query Match 11.8%; Score 442; DB 11; Length 1091;
 Best Local Similarity 23.2%; Pred. No. 6, 9e-25;
 Matches 169; Conservative 106; Mismatches 253; Indels 198; Gaps 22;

QY 5 VADLLAW-----AGATATVVPVWVPCPPQCAC-----QI 37
 DB 14 LAPRLILLILLILLOWPESAGAGARP-----RAPCAACTCAGNSLDCSGRLATLPRL 68
 QY 38 RPYVTPRS---SYREATYVD-----CNDLFLTVAPPALPA--GTOTLLLOS 78
 DB 69 PSW--TRSLNLSYNRLEISDAFEDLTNIQEVYLNNSNLTPIPSLGTASISGVVSFTDQ 126
 QY 79 NSIVRDOSELIYANLTLELDSONSFSDARCDFFALPOLLSLHLEENOLRLDHSFA 138
 DB 127 NKLISVDSQKSYSLSEVLDLSSNNITIRSSCPPNGRLRIELINLAKRISITLISGAFD 186
 QY 139 GLA-STQELYLNNHQLYRIAPAFSGLSNLRLHNSNLRAIDSRWEPMLNLEIIMG 197
 DB 187 GLSRSLTLRLSKNRITQLPVKAFL-IPRLTOLDINRRNIRLIEGLTGGDLSLEVLRQ 245
 QY 198 GKNVDAIILDMNRPRLANLSYLVLAGNMLEIDYALEGLQSLESISFDNQLARVPR--- 254
 DB 246 RNNISRLTGAFWGLSKMVLHLEYNLSVEVNSGLYGLTALHOLHLSNNSISRIORGW 305
 QY 255 -----RALBOVPGKFLDKLNKNPQVGVGDPAANMLHLELGIN 293
 DB 306 STCQKHEILISFNNTLRDEESLAEISLTLRLSHNAISITAGATGKGLSLVLDID 365
 QY 294 NNEELVSD--KFAVLNPELTKLDTNNPRLSFIHPRAFHILPOMETLMLNNALSLH 351
 DB 366 HNEISGTLIDTSGAFTGIDNLSKLTLPFGN-KIKSVAKRAFSQLESLEHNLNGENAIRSVQ 424

QY 352 QQVTESLPNLQEVGLHGNPIRDCVIRW-----ANAT-----GTRVRFI 390
 DB 425 FDAFAKMNKLKELYISSSFLDCQKWLPPMLGRMLQAFVATCAHPESLKQSIFSV 484
 QY 391 EPQSLCAEPP-----DL 403
 DB 485 LPDSFVCDFFRPQIITQPEPTMAVVGKDIRFTCSAASSSSPMTFAMKDNVLANAD 544
 QY 404 QR-----LPREVFP-----REKTDHCLPLIS 425
 DB 545 ENFAHVRADQGEVMEYTTILRLRHVTGHEGRYQCIITNHEGSTYSIKARLTVAVLP--S 602
 QY 426 PRSEPPSIQVAGSGSMVLHCRALAEPEPELYVTTPAGILRTPAHAGRRYRYP-EGTLEL 484
 DB 603 FTKIPHDIAIRGTGTLARLECAATGHPNPQIAMQKDGTFD-PAARERRMHVMPDDVFEI 661
 QY 485 RRYVAEEGLTCAVQNIYVG--ADTKYVYVVGALLQPGNDE----GQGLRLRVQDTHP 538
 DB 662 TDVKIDMGVYSCTPAONSAGSVSANATLTLETPLAVPLEDRVVTGVTAFQCKATGS 721
 QY 539 YHLLSW 546
 DB 722 PTPRTWL 729
 RESULT 15
 ID 09HCJ2 PRELIMINARY; PRT; 640 AA.
 AC 09HCJ2;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)
 DE K1A1580 PROTEIN (FRAGMENT).
 GN K1A1580.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20450683; PubMed=10997877;
 RA Nagase T., Kikuno R., Nakayama M., Hirosewa M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes.
 XVIII. The complete sequences of 100 new cDNA clones from brain which
 code for large proteins in vitro.";
 RL DNA Res. 7:273-281(2000).
 DR EMBL: AB046800; BAB13406.1; -
 DR HSSP: P22888; ITUM.
 DR InterPro: IPR003599; Ig.
 DR InterPro: IPR003598; Ig_C2.
 DR InterPro: IPR003600; Ig_Like.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR000483; LRR_Cterm.
 DR InterPro: IPR000372; LRR_Nterm.
 DR InterPro: IPR003592; LRR_out.
 DR Pfam: PF00047; Ig_1.
 DR Pfam: PF00560; LRR; 9.
 DR Pfam: PF01463; LRRC1.
 DR Pfam: PF01462; LRRC1.
 DR PRINTS; PR00019; LEURICHART.
 DR SMART: SM00409; Ig; 1.
 DR SMART: SM00408; IgC2; 1.
 DR SMART: SM00410; Ig_Like; 1.
 DR SMART: SM00370; LRR; 6.
 DR SMART: SM00082; LRRC1; 1.
 DR SMART: SM00013; LRRC1; 1.
 DR SMART: SM00369; LRR_Typ; 8.
 KW Immunoglobulin domain.
 FT NON TER 1
 SQ SEQUENCE 640 AA; 71949 MW; 6058974872636838 CRC64;

Query Match 11.7%; Score 438.5; DB 4; Length 640;
Best Local Similarity 25.5%; Pred. No. 5.8e-25;
Matches 148; Conservative 84; Mismatches 231; Indels 117; Gaps 12;

```
OY 2 RLVAFLDLAVAGATATVPVPHVHPQACQIRPWYTPRSSYREATTVDCNDLELT 61
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 20 RALFDPPLVLLAQQLLVAGLVRAQCPSVCS-----NPFSKVICVRKNLR 68
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 62 AVPPALPAGQOTLLQNSIVRDQSEGLATLUTELDLSQNSFSDARDCDFHALPQLLS 121
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 69 EVPQISTNTRLLNLHNSQIQLIKVNSFKRLHLEITQLSRNRIETIEGAENLANMT 128
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 122 LHLBNQTLRLDEHSFAGLASLOELYLNHQLYRIAPRAFSGLSNLLRLNLSNLLRAID 181
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 129 LELFDNRLLTTPNGAFVYLSKRLKELMLRNPNIESIPSYAFNRIPSLRDLDE----- 181
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 182 SRWEEMLPNLEIMIGNKVDAILDMNFRPLANKSLVLAGMNLREISDYALEGLQSLSES 241
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 182 -----LKRLSYISEGA-----FEGLSNLRYLNLAMCNLREIPN----- 214
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 242 LSFYDNLARVPRRALQCVGKFLDLKNPQROVGPDPANMLHLKELGNNMELVSI 301
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 215 -----LPLIKDELDELDSGNHLSAIRGSPQGLMHLOKLMIGSQ----- 254
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 302 DKFALVNLPELTKLDITRNPRLSFIHPRAFHLLPOMETLMLNNALSALHQTVESLPNL 361
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 255 -----IOVIERNAFDNLOSLVEINLAHNNLTLLPHDLFTPLHL 293
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 362 QEVGLHGNPIRCDCVIRMANATGTRVRIEPOSTLCA---EPPDLQRLPREVPREMT 417
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 294 ERIHHNNPWNQNCDIWLSS--WIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFT 350
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 418 DHCLPLISPRSPSLQVASESMVLCGRALAEPEPEIYWTTPAGLRILTPAHNGRRYRY 477
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 351 CYAPYIPEP--PADLVNTEGMAELKCRG--STLSISVSWITPNGVTMGAYKVRIRAYL 406
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 478 PEGTLELRRTVAEEAGLYTCVAQNLVGADTKTVSVVVGRA-----LLQPGR 523
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 407 SDGTINFTNVTYQDTGMTCVNSVSGNTTASATLANTATTPPSYFSTVETMEPSQ 466
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 524 DEGGGLELRVOETHPRYHLLSWTPPNTVSTNLTWSASS 563
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 467 DEARTDNNVGPTP---VVDMEF--TNVTTISLTPOSTRS 500
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: August 31, 2002, 14:40:33
Job time: 260 sec

